A Dissertation for the Degree of Doctor of Philosophy

Structure and Dynamics of the S100A5-RAGE complex and anti-CRISPR AcrIIA4 probed by NMR spectroscopy

핵자기 공명 분광학을 이용한 S100-RAGE 복합체와 anti-CRISPR AcrIIA4 의 구조 및 동력학 연구

February, 2018

Biomodulation Major
Department of Agricultural Biotechnology
Seoul National University

Iktae Kim

ABSTRACT

Structure and Dynamics of the S100A5-RAGE complex and anti-CRISPR AcrIIA4 probed by NMR spectroscopy

Iktae Kim

Biomodulation Major

Department of Agricultural Biotechnology

Seoul National University

Distinct protein-protein interactions were investigated using NMR spectroscopy as well biophysical characterization tools such as calorimetry and size exclusion chromatography. In Chapter I, calcium binding of S100A5 and the interaction between S100A5 and receptor for advanced glycation end product (RAGE) were investigated using calorimetry and NMR spectroscopy. S100A5 exhibits a sequential manner to bind calcium ions using the C-terminal and N-terminal calcium binding EF-hands with the equilibrium dissociation constants (K_D) of 1.3 μM and 3.5 μM, respectively. S100A5 interacts with the V domain of RAGE (RAGEv) to form a

heterotrimeric complex with K_D of 5.9 μM in a calcium-dependent manner, which is a distinct stoichiometry from the S100 family proteins. Chemical shift perturbation data from NMR titration experiments indicates that S100A5 employs the periphery of the dimer interface to interact with RAGEv. Distinct binding mode and stoichiometry of against different S100 family proteins could be important to modulate diverse RAGE signaling.

In Chapter II, solution structure and dynamics of AcrIIA4 that inhibits *Streptococcus pyogenes* Cas9 (*Spy*Cas9) widely used for genome editing were investigated. The structure of AcrIIA4 is a compact monomeric αβββαα fold consisting of three α helices flanked against three antiparallel β strands and short 3_{10} helix. The backbone dynamics of AcrIIA4 reveals that distinct backbone dynamics in fast and slow timescales at loop regions is related to interaction surface for *Spy*Cas9. Furthermore, AcrIIA4 binds to apo-*Spy*Cas9 with K_D ~4.8 μM and ~0.6 nM for AcrIIA4 binding to sgRNA-bound *Spy*Cas9, which infers that a binary complex between AcrIIA4 and *Spy*Cas9 can associate with sgRNA to form a tight ternary complex, avoiding competition with the DNA substrate for *Spy*Cas9 binding. This studies provide insights into anti-CRISPR-mediated inhibition mechanisms for disabling the *Spy*Cas9, thereby broadening CRISPR-Cas regulatory tools for genome editing.

Keyword: AcrIIA4, Anti-CRISPR, Cas9, NMR spectroscopy, RAGE, S100A5, Structure and dynamics

Student Number: 2011-24110

CONTENTS

LIST OF FIGURES

CHAPTER I

CHAPTER II

LIST OF TABLES

CHAPTER I

CHAPTER II

ABBREVIATION

BME	β-Mercaptoethanol
CBCACONH	Correlation spectroscopy for H^N–N^H–$C\alpha^i$–$C\beta^i$
CSP	Chemical shift perturbation
DTT	dithiothreitol
EDTA	Ethlyenediaminetetraacetic acid
HBHACONH	Correlation spectroscopy for H^N–N^H–H^α–H^β
HNCACB	Correlation spectroscopy for $C\alpha^{i-1}$–$C\beta^{i-1}$–H^N–N^H–$C\alpha^i$–$C\beta^i$
HSQC	Heteronuclear single quantum coherence spectroscopy
IPTG	Isopropyl β–D–1–thiogalactopyranoside
ITC	Isothermal titration calorimetry
MALS	Multi-angle light scattering
MOPS	3-(N-morpholino) propanesulfonic acid
NMR	Nuclear magnetic resonance
NOESY	Nuclear Overhauser effect spectroscopy
OD	Optical density
PMSF	Phenylmethanesulfonyl fluoride
RDC	Residual dipolar coupling
SDS–PAGE	Sodium dodecyl sulfate–polyacrylamide gel electrophoresis
TEV	Tobacco Etch Virus
TOCSY	Total correlation spectroscopy

CHAPTER I

Biophysical characterization of calcium-binding of S100A5 and calcium induced S100A5-RAGE complex

1. Abstract

Receptor for advanced glycation end product (RAGE) as a pattern recognition receptor plays an important role, such as ligand recognition, binding and signal transduction on cell surface. A variety set of ligand such as AGEs, Aβ peptide, HMGB1 and S100 proteins binds to RAGE to trigger downstream intracellular signaling. S100A5 is a small calcium binding protein and represents a major ligand of RAGE. Here I have characterized calcium binding of S100A5 and the interaction between S100A5 and RAGE using calorimetry and NMR spectroscopy. S100A5 exhibits a sequential manner to binds calcium ions using the C-terminal and N-terminal calcium binding EF-hands with the equilibrium dissociation constants (K_D) of 1.3 μM and 3.5 μM, respectively. S100A5 interacts with the V domain of RAGE (RAGEv) to form a heterotrimeric complex with K_D of 5.9 μM in a calcium-dependent manner, which is a distinct stoichiometry from the S100 family proteins. Chemical shift perturbation data from NMR titration experiments indicates that S100A5 employs the periphery of the dimer interface to interact with RAGEv. Distinct binding mode and stoichiometry of against different S100 family proteins could be important to modulate diverse RAGE signaling.

2. Introduction

2. 1. Receptor for Advanced Glycation End product (RAGE)

The receptor for advanced glycation end products (RAGE) is a pattern recognition receptor which binds a variety set of ligand on cell surface. It consists of three extracellular domains (V-type, C1-type, and C2-type), a transmembrane helix domain and a short cytosolic domain[1-3]. RAGE plays an important role for the signaling pathway in chronic inflammatory disease[4,5], diabetes complications[6], Alzheimer's disease[7], and cancer[8]. Various ligands including AGE products, nucleic acids, Aβ peptide, HMGB1, and S100 proteins binds RAGE to trigger intracellular signaling for diverse physiological and pathological effects[9-13] (Fig. 1). Expression of RAGE varies depending on cell types and ligand concentrations[14,15], and initiates multiple signaling cascades, such as the MAPK, JNK, and Cdc42/Rac pathways, leading to activation of transcription factors AP-1 and NF-κB[16,17].

2. 2. S100A5

S100 proteins are small Ca^{2+}-binding proteins with two EF-hand motifs. More than 20 members of S100 proteins have been identified in human so far, and they exist mostly as a homo- or heterodimer[18,19]. S100A5, called S100D, has first isolated from bovine brain and is highly expressed in the cells and fibers of the hippocampus and recurrent grand I meningioma[20-22]. It is also expressed in restricted areas of the human kidney and upregulated in bladder cancer[23]. An S100A5 dimer accommodates four Ca^{2+} ions with strong positive cooperativity, and exhibits 20-100

fold higher Ca^{2+}-binding affinity than other S100 proteins[20,24]. S100A5 also binds to Zn^{2+} and Cu^{2+} ions, and it is implicated that S100A5 may protect other proteins from Cu^{2+}-induced oxidative stress similarly as S100B[20,25,26]. Solution structures of S100A5 has been determined using NMR spectroscopy both in apo state and the holo state, where the backbone folds were largely similar with those of other S100 family protein[24]. S100 proteins are known as one of the main ligands for the RAGE, and participate in diverse RAGE signaling pathway[27], in spite of high structural homology between S100 family proteins (Fig 1 and 2). Interaction of RAGE with different S100 proteins has been actively investigated by X-ray crystallography and NMR spectroscopy[28-31]. It has been reported that S100B, S100A6, S100P and AGEs induces homo-dimerization of extracellular domain, which amplifies RAGE-mediated signal transduction, in spite of high structural homology between S100 family proteins[32-34].

In this study, the Ca^{2+}-binding of S100A5 was characterized using isothermal titration calorimetry (ITC), and the complex formation between S100A5 and RAGE was examined using size exclusion chromatography, ITC, and NMR spectroscopy. The binding thermodynamics of individual EF-hands in S100A5 indicated that S100A5 binds to calcium ions in a sequential manner in that C-terminal EF-hands are saturated and followed by N-terminal EF-hands. The characterized complex formation revealed that S100A5 dimer binds to one molecule of the V domain of RAGE (RAGEv), forming a heterotrimer that is distinct among the S100 protein family.

Figure 1. Illustration of RAGE signaling induced by various ligands

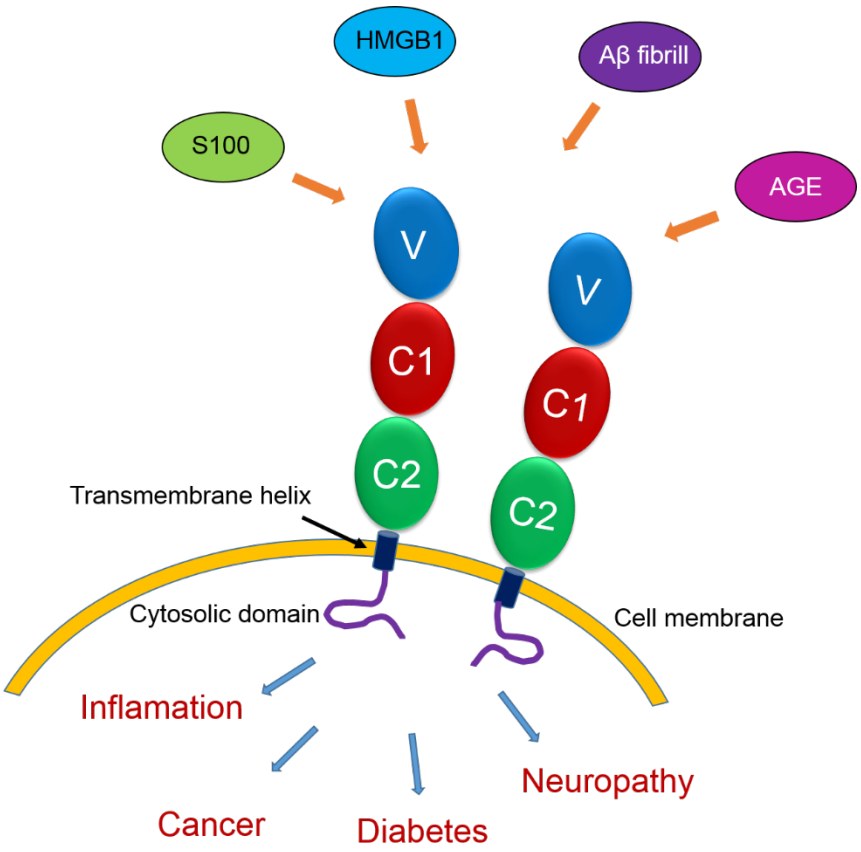
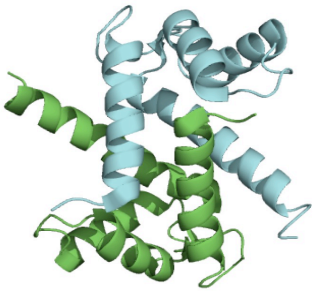
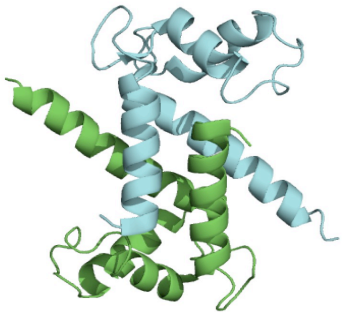

HMGB1
Aβ fibrill
S100
AGE
V
C1
C2
V
C1
C2
Transmembrane helix
Cytosolic domain
Cell membrane
Inflamation
Cancer
Diabetes
Neuropathy

Figure 2. Structural homology between S100 family proteins

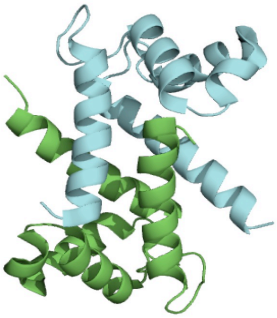
In the dimeric S100 family proteins as ribbon diagram, one subunit of the dimer is colored in cyan and the others in green. In spite of the high structural homology, it has been suggested that binding interfaces for RAGE are not conserved.



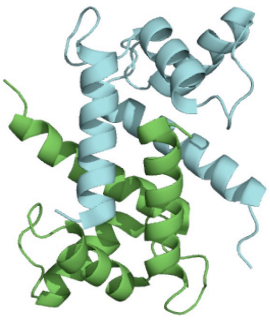
S100A5



S100P



S100A6



S100B

3. Materials and Methods

3. 1. Cloning

Each gene of S100A5 (residues 1–92) and RAGEv (residues 23–125) were cloned into a pET32a vector (Novagen) with thioredoxin tags, respectively. S100A5 C43S (S100A5m) was mutated using the QuickChange kit (Agilent Technologies), and the constructs were verified by DNA sequencing.

3. 2. Protein expression

3. 2. 1. Protein expression - Luria bertani medium

The expression of the S100A5, S100A5m and RAGEv was performed using the expression vector pET32a in *Escherichia coli* strain BL21star (DE3) (Invitrogen) and Origami B (DE3) (Novagen), respectively, using LB medium (Table 1). The cells with expression vector were cultured at 37°C to an OD_{600} of 0.6–0.8, and the protein overexpression was induced by adding IPTG to a final concentration of 1 mM, followed by induction at 37°C and harvested at OD_{600} of 1.2–1.6 by centrifugation. The cell separated from medium was stored at -80^{o}C

3. 2. 2. Protein expression - Isotope labeling medium

[^{15}N]/[^{13}C/^{15}N]-labeled proteins are commonly used for increasing the sensitivity of NMR experiments. The overexpression vectors were transformed into *Escherichia. coli* strain BL21star (DE3) (Invitrogen) or Origami B (DE3) (Novagen),

and the cells were grown in 2 mL LB medium for 2 h to inoculate to the M9 minimal media (with $^{15}NH_4Cl$ and/or $^{13}C_6$-glucose as the sole nitrogen or carbon source, respectively) (Table 1). The Cells were grown at 37°C until an OD_{600} of 0.6–0.8 and induced with 1 mM IPTG at 37°C for approximately 8 h or overnight grown at 17°C, harvested by centrifugation and then stored at -80^{o}C.

Table 1. The composition of media.

Medium	Composition (per Liter)
LB medium	10 g Tryptone, 5 g Yeast extract, 10 g NaCl
M9 minimal medium (^{15}N or $^{13}C/^{15}N$)	10 g K_2HPO_4, 13 g K_2HPO_4, 9 g Na_2HPO_4, 2.4 g K_2PO_4, 1 g $^{15}NH_4Cl$, 2 g U–$^{13}C6$ Glucose (or 5 g Glucose), 10 mM $MgCl_2$, 0.1 mM Thiamine, 1 x Trace element, 0.2 mM $CaCl_2$, 50 mg Carbenicillin

3. 3. Purification

3. 3. 1. S100A5 and S100A5 C43S (S100A5m)

The cells were harvested by centrifugation, and resuspended in buffer 20 mM Tris–HCl (pH 7.4), 200 mM NaCl, 5 mM BME, 1 mM phenylmethanesulfonyl fluoride (PMSF, Sigma, P7626), and disrupted by emulsiflex (Avestin, Canada). The lysate was centrifuged at 24,000g for 30 min at 4°C. The supernatant was loaded on a HisTrap HP column (5 ml, GE Healthcare) column equilibrated with binding buffer 20 mM Tris–HCl (pH 7.4), 200 mM NaCl, 5 mM BME and eluted with a gradient 100 ml of 500 mM Imidazole. The solution containing protein was fractionated and His_6-tag was removed using TEV protease while dialyzed against 50 mM Tris-HCl (pH 8.0), 50 mM NaCl, 5 mM BME for overnight. The digestion reaction was loaded onto HisTrap column. The flow-through was collected and loaded onto a Hiload Superdex 75 column (GE Healthcare) pre-equilibrated with 20 mM Tris-HCl, (pH 7.4), 200 mM NaCl, 5 mM BME and finally were purified by Mono Q column (GE Healthcare) using a gradient of 1 M NaCl. The fractionated S100A5 or S100A5m was pooled with 20 mM EDTA, 20 mM Tris-HCl, (pH 7.4), 200 mM NaCl, 5 mM BME and washed with 20 mM Tris-HCl, (pH 7.4), 200 mM NaCl, 2 mM dithiothreitol (DTT). The purified S100A5 and S100A5m were analyzed by SDS–polyacrylamide gel electrophoresis to confirm sample mass and sample purity (Figure 3).

3. 3. 2. RAGEv

The cell containing the RAGEv protein was resuspended in buffer 50 ml of 20 mM Tris, pH 7.4, 200 mM NaCl, and 1 mM PMSF. The suspension was lysed by

three passages through a emulsiflex and centrifuged at 20,000 g for 30 min at 4°C. The supernatant was loaded on a HisTrap HP column (5 ml, GE Healthcare) column equilibrated with binding buffer 20 mM Tris–HCl (pH 7.4), 200 mM NaCl and eluted with a gradient 100 ml of 500 mM Imidazole. The solution containing protein was fractionated and His_6-tag was removed using Thrombin protease (Sigma-Aldrich, T6885-1KU) while dialyzed against 20 mM Tris-HCl (pH 7.4), 200 mM NaCl, 5 mM for overnight. The digestion reaction was loaded onto HisTrap column. The flow-through was collected and loaded onto a Hiload Superdex 75 column (GE Healthcare) pre-equilibrated with 20 mM Tris-HCl, (pH 7.4), 200 mM NaCl and finally were purified by Mono S column (GE Healthcare) using a gradient of 1 M NaCl. The purified RAGEv were analyzed by SDS–polyacrylamide gel electrophoresis to confirm sample mass and sample purity (Figure 4).

Figure 3. Purification procedure of S100A5 and S100A5m

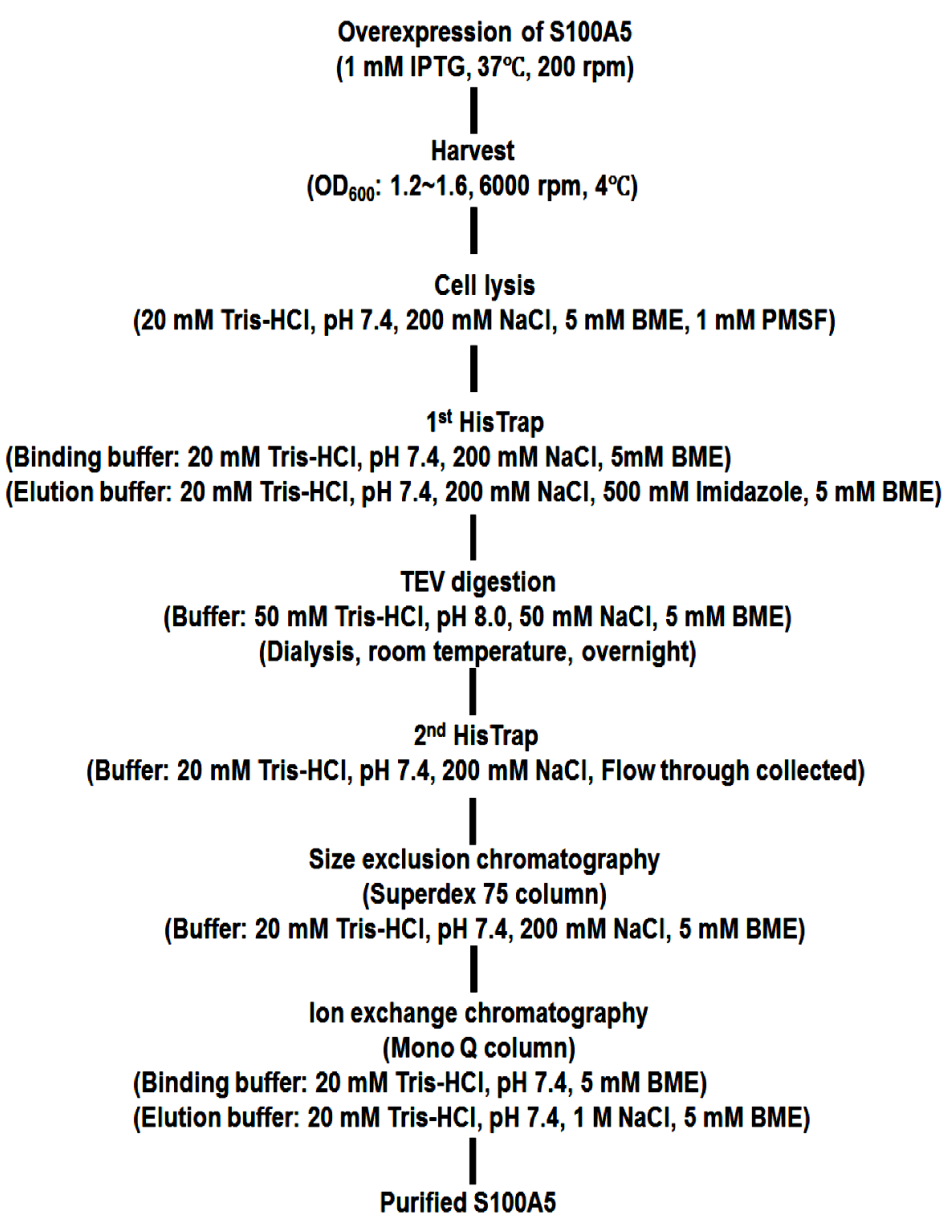
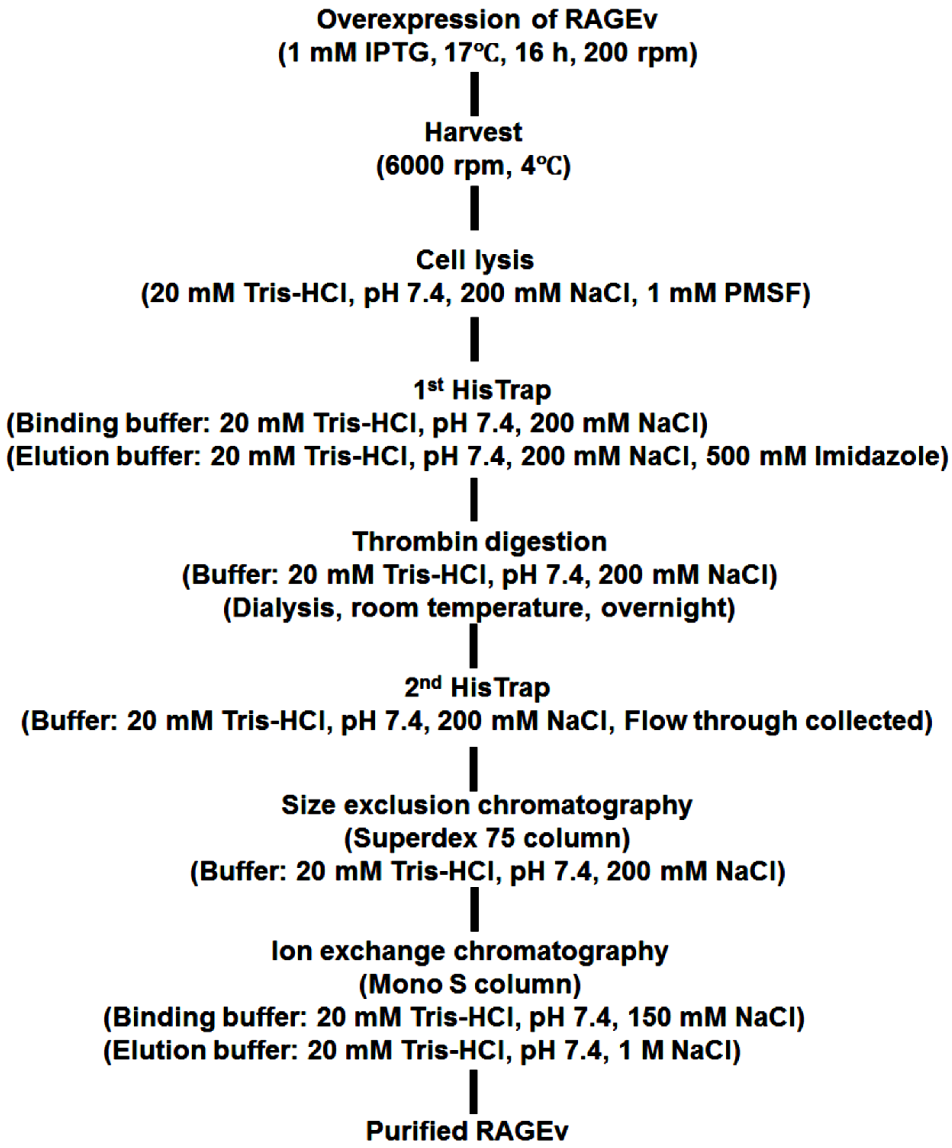

Overexpression of S100A5
(1 mM IPTG, 37℃, 200 rpm)
Harvest
(OD_{600}: 1.2~1.6, 6000 rpm, 4℃)
Cell lysis
(20 mM Tris-HCl, pH 7.4, 200 mM NaCl, 5 mM BME, 1 mM PMSF)
1st HisTrap
(Binding buffer: 20 mM Tris-HCl, pH 7.4, 200 mM NaCl, 5mM BME)
(Elution buffer: 20 mM Tris-HCl, pH 7.4, 200 mM NaCl, 500 mM Imidazole, 5 mM BME)
TEV digestion
(Buffer: 50 mM Tris-HCl, pH 8.0, 50 mM NaCl, 5 mM BME)
(Dialysis, room temperature, overnight)
2nd HisTrap
(Buffer: 20 mM Tris-HCl, pH 7.4, 200 mM NaCl, Flow through collected)
Size exclusion chromatography
(Superdex 75 column)
(Buffer: 20 mM Tris-HCl, pH 7.4, 200 mM NaCl, 5 mM BME)
Ion exchange chromatography
(Mono Q column)
(Binding buffer: 20 mM Tris-HCl, pH 7.4, 5 mM BME)
(Elution buffer: 20 mM Tris-HCl, pH 7.4, 1 M NaCl, 5 mM BME)
Purified S100A5

Figure 4. Purification procedure of RAGEv.


Overexpression of RAGEv
(1 mM IPTG, 17℃, 16 h, 200 rpm)
Harvest
(6000 rpm, 4℃)
Cell lysis
(20 mM Tris-HCl, pH 7.4, 200 mM NaCl, 1 mM PMSF)
1st HisTrap
(Binding buffer: 20 mM Tris-HCl, pH 7.4, 200 mM NaCl)
(Elution buffer: 20 mM Tris-HCl, pH 7.4, 200 mM NaCl, 500 mM Imidazole)
Thrombin digestion
(Buffer: 20 mM Tris-HCl, pH 7.4, 200 mM NaCl)
(Dialysis, room temperature, overnight)
2nd HisTrap
(Buffer: 20 mM Tris-HCl, pH 7.4, 200 mM NaCl, Flow through collected)
Size exclusion chromatography
(Superdex 75 column)
(Buffer: 20 mM Tris-HCl, pH 7.4, 200 mM NaCl)
Ion exchange chromatography
(Mono S column)
(Binding buffer: 20 mM Tris-HCl, pH 7.4, 150 mM NaCl)
(Elution buffer: 20 mM Tris-HCl, pH 7.4, 1 M NaCl)
Purified RAGEv

3. 4. Size exclusion chromatography and MALS analysis

Size exclusion chromatography (SEC) analysis was performed on Superdex 75 10/300 GL column (GE Healthcare) at 25°C. The column was equilibrated with 20 mM Tris-HCl, (pH 7.4), 100 mM NaCl, 10 mM $CaCl_2$ buffer. 1.6 mM of S100A5 and 0.8 mM RAGEv were loaded onto the column individually or together. Eluted samples were confirmed by SDS-PAGE. SEC-MALS experiments were performed using an Agilent 1200 HPLC system (Agilent Technologies) combined with a Wyatt DAWN HELEOS-II MALS instrument and a Wyatt OPtilab rEX differential refractometer (Wyatt). For chromatographic separation, WTC-030S5 column (Wyatt) was used at a flow rate of 0.5 ml/min in 20 mM Tris-HCl, (pH 7.4), 100 mM NaCl, 10 mM $CaCl_2$. Injection concentration of S100A5m, RAGEv and S100A5m/RAGEv complex were 1 mM, 1.5 mM and 1 mM/1.5 mM, respectively. The equipment was pre-calibrated with 1 mg/ml bovine serum albumin (monomer ~66 kDa, and dimer ~133 kDa) and the results were normalized and analyzed using ASTRA software (Wyatt)

3. 5. NMR experiment

The NMR samples contained 1 mM ^{13}C,^{15}N-S100A5m or 0.8 mM ^{13}C,^{15}N-RAGEv in 20 mM Tris-HCl, (pH 7.4), 100 mM NaCl. NMR spectra were collected on Bruker AVANCE 600 MHz and 900 MHz spectrometers equipped with a x,y,z-shielded gradient triple resonance cryoprobe. Sequential and backbone assignments of ^{1}H,^{15}N and ^{13}C resonances were achieved by three-dimensional triple resonance through-bond scalar correlation CBCACONH and HNCACB experiments at 25°C. For NMR titration, ^{1}H-^{15}N HSQC spectra recorded for 0.4 mM of S100A5m or RAGEv titrated

with the partner protein and changes in the backbone amide chemical resonances were monitored. NMR spectra were processed using the NMRPipe[35] program and analyzed PIPP[36]and the NMRView[37] programs.

3. 6. Isothermal titration calorimetry (ITC)

ITC experiments were performed at at 25°C using iTC200 calorimeter (GE Healthcare). To characterize Ca^{2+}-affinity of S100A5, 0.1 mM S100A5 was titrated with 2 mM $CaCl_2$ in 30 mM MOPS, (pH 7.4), 100 mM NaCl, 5 mM BME. Twenty consecutive 2 μL aliquots of proteins were titrated into the cell. The duration of each injection was 4 s, and injections were made at intervals of 180 s. For titration between S100A5m and RAGEv, 0.2 mM S100A5m was titrated with 1 mM RAGEv in 20 mM Tris-HCl, (pH 7.4), 100 mM NaCl, 5 mM $CaCl_2$. The heats associated with dilution of the substrates were subtracted from the measured heats of binding. ITC data were analyzed with the Origin version 7.0 program provided with the instrument.

4. Results

4.1. Thermodynamics of calcium binding events in S100A5

In general, S100 family proteins have two calcium binding site, C-terminal canonical EF-hand and an N-terminal S100-specific EF-hand, and the C-terminal EF-hand exhibits a stronger Ca^{2+} binding affinity. Because the S100 family proteins generally form a dimer, four calcium ions can be accommodated in a S100 homodimer, which induced a large conformational change of S100 protein. I measured the Ca^{2+} binding affinity of S100A5 using ITC. The binding isotherm was fit into a sequential binding model, such that four Ca^{2+} ions bound to an S100A5 dimer with two K_D values, K_{D1} of 1.3 μM and K_{D2} of 3.5 μM (Fig. 5A, and Table 2). Here, each K_D value represents the binding event of two Ca2+ ions, such that there are two sets of two equivalent binding sites. Based on the sequential binding model, Ca^{2+} ions can bind to S100A5 in two different ways. First, Ca^{2+} ions may bind to each C-terminal EF-hand in the S100A5 dimer, and then proceed to bind to the N-terminal EF hands. Second, Ca^{2+} ions may fill up one S100A5 subunit in the dimer, and then proceed to the next subunit. I note that S100A5 undergoes a significant conformation rearrangement in the C-terminal EF-hand upon Ca^{2+}-binding (Figs. 5 B and C)[24], so that Ca^{2+}-binding to one subunit would results in an asymmetric dimer that can be identified from the $^{1}H–^{15}N$ HSQC spectra of S100A5 upon Ca^{2+} titration. I carried out a series of NMR titration experiments with S100A5 and Ca^{2+} at various temperatures (Fig 6) and there did not appear extra backbone amide resonances

attributed to the asymmetric dimer. I thus interpret the measured K_{D1} and K_{D2} values as the Ca^{2+}-binding constants at C-terminal and N-terminal EF-hands of S100A5, respectively. K_{D1} represents the intrinsic Ca^{2+}-binding of C-terminal EF-hand in the apo state of S100A5, and K_{D2} represents the Ca^{2+}-binding of N-terminal EF-hand when Ca^{2+} is loaded in C-terminal EF-hand. Previous study using flow dialysis showed that Ca^{2+} ions bind to S100A5 in a highly cooperative manner[20]. I have quantitatively determined Ca^{2+}-affinity of each EF-hand, such that Ca^{2+}-binding of the C-terminal EF-hand is ~3 times stronger than that of the N-terminal EF-hand. Note that the intrinsic Ca^{2+}-affinity of the N-terminal EF-hand in the apo-state could be much weaker, considering the cooperative nature of Ca^{2+}-binding of S100A5.

The binding enthalpy (ΔH) and entropy ($-T\Delta S$) of Ca^{2+}-binding were very different between the C-terminal and N-terminal EF-hands. Ca^{2+}-binding at the C-terminal EF-hand was mostly driven by a favorable entropic contribution with $-T\Delta S$ of -8.0 ± 0.1 kcal·mol^{-1}, and little enthalpic contribution. On the other hand, Ca^{2+}-binding at the N-terminal EF-hand exhibited a considerably larger enthalpic contribution with ΔH of -3.2 ± 0.1 kcal·mol^{-1} and $-T\Delta S$ of -4.2 ± 0.1 kcal·mol^{-1}. The large difference in the thermodynamic parameters likely originated from the Ca^{2+}-induced conformational change. The conformational rearrangement is manifest mainly in the C-terminal EF-hand, which involves a large reorientation of helix III and the elongation of helix IV, whereas the N-terminal EF-hand exhibits only a minor change. I speculate that the large entropic contribution in the Ca^{2+}-binding to the C-terminal EF-hand of apo S100A5 implicates the presence of a conformational ensemble exchanging between Ca^{2+}-bound and Ca^{2+}-free structures. Ca^{2+}-binding to the N-terminal EF-hand of half-saturated S100A5 forms a fully saturated holo state,

and stabilizes the Ca^{2+}-bound structure, which accounts for the favorable enthalpic contribution.

Figure 5. Thermodynamics of Ca^{2+}-binding of S100A5.

Raw ITC data (A, top panel) and integrated heats of injection (A, bottom panel) for he titration of Ca^{2+} and S100A5. The squares in the bottom panels are the experimental data, and the lines represent the least-squares best-fit curve derived for a sequential binding model. The structures of S100A5 are shown in a ribbon diagram for (B) the apo state (PDB code 2kax) and (C) the holo state (PDB code 2kay). Ca^{2+} ions are shown in orange spheres. Only one subunit of the dimer is shown for visual clarity.

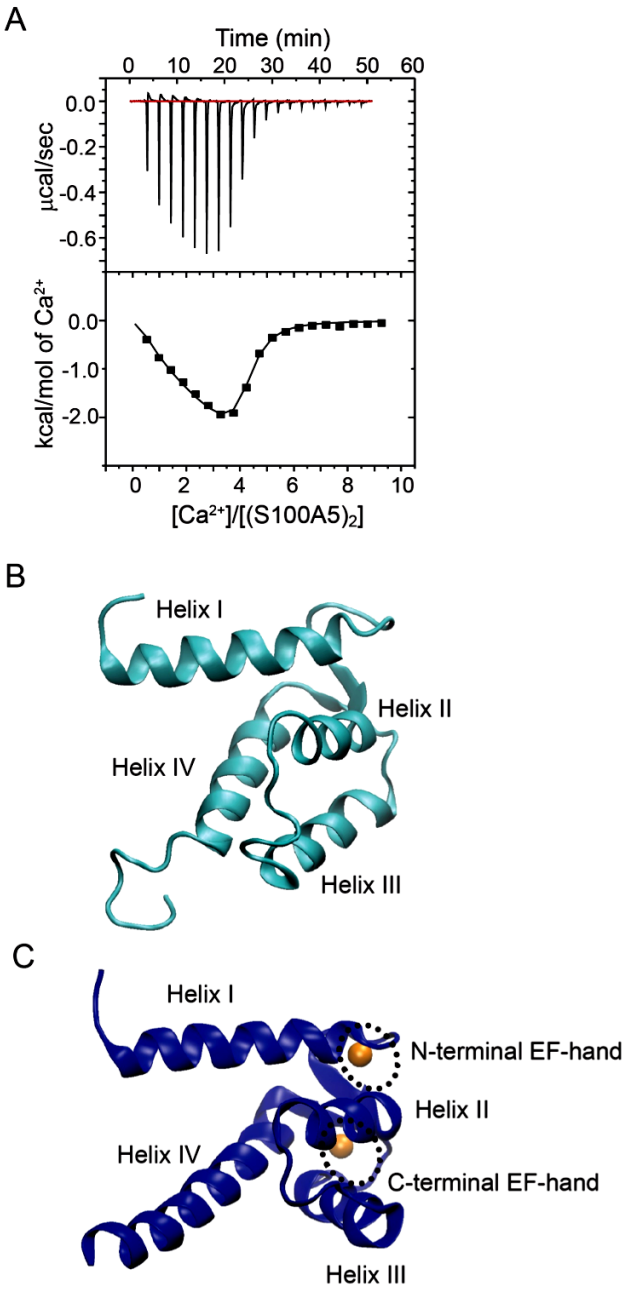

A
Time (min)
0 10 20 30 40 50 60
0.0
-0.2
-0.4
-0.6
µcal/sec
0.0
-1.0
-2.0
kcal/mol of Ca2+
0 2 4 6 8 10
[Ca2+]/[(S100A5)2]
B
Helix I
Helix II
Helix IV
Helix III
C
Helix I
N-terminal EF-hand
Helix II
Helix IV
C-terminal EF-hand
Helix III

Figure 6. Superimposed ^{1}H-^{15}N spectra from NMR titration of ^{15}N-S100A5 with Ca^{2+}.

0.4 mM of S100A5 was stoichiometrically titrated with Ca^{2+} at 25℃ according to the molar ratio that is color-coded as the bar graph in the spectra.

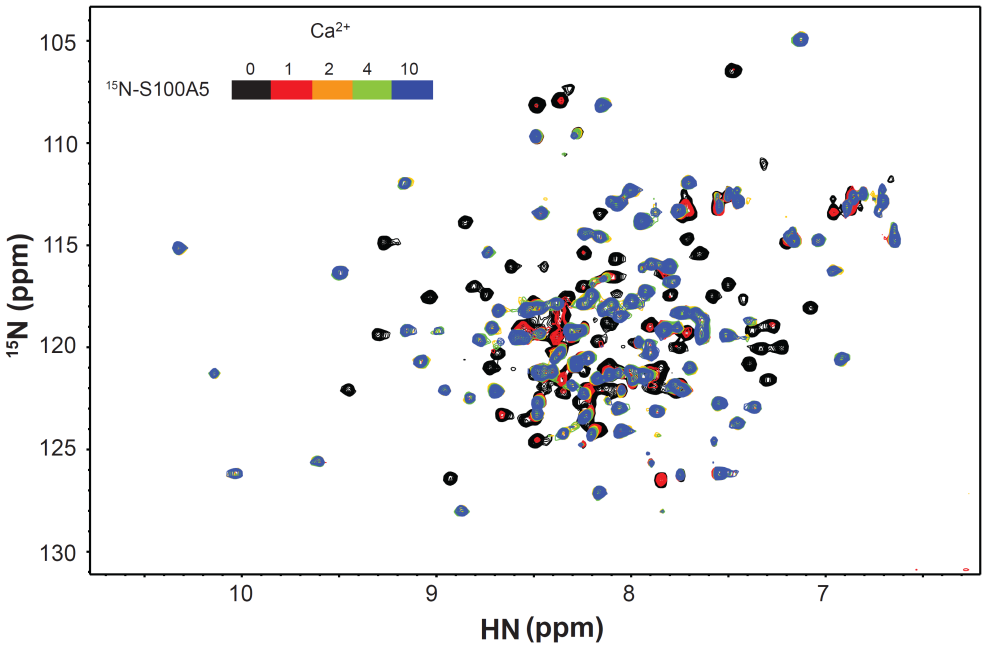

Ca2+
0 1 2 4 10
15N-S100A5
105
110
115
120
125
130
15N (ppm)
10
9
8
7
HN (ppm)

Table 2. Equilibrium dissociation constants (K_D), free energy changes (ΔG), enthalpy changes (ΔH), and entropy changes ($T\Delta S$) for the Ca^{2+}-binding of S100A5. The binding isotherm was fit into a sequential binding model, such that four Ca^{2+} ions bound to an S100A5 dimer.

	K_D (μM)	ΔG (kcal/mol)	ΔH (kcal/mol)	$T\Delta S$ (kcal/mol)
S100A5:Ca^{2+}	1.3 ± 0.2 μM	-8.0 ± 0.1	0.0 ± 0.1	8.0 ± 0.1
	3.5 ± 0.5 μM	-7.4 ± 0.1	-3.2 ± 0.1	-4.2 ± 0.1

4.2. Resonance assignment of S100A5m

NMR spectra of Calcium bound S100A5 and S100A5m

RAGEv contains a disulfide bond between Cys38 and Cys99 for proper folding, which prevents a use of reducing agent. S100A5 has two free cysteine, Cys43 and Cys80. The solution structure of S100A5 indicated that Cys43 was exposed to the solvent, while Cys80 was buried at the dimer interface. Cys43 of S100A5 was mutated to Ser43 and the mutant S100A5 (S100A5m) remained as a monomer in the absence of reducing agents in the SDS-PAGE. S100A5m exhibited a similar 2D ^{1}H-^{15}N HSQC spectra as wild-type S100A5 (Fig. 7) Thus, I employed S100A5m to characterize the interaction between S100A5 and RAGEv,

NMR backbone assignment of S100A5m

Backbone assignment of S100A5m consisting of 92 amino acids with molecular weight of 10.8 kDa has carried out using a suite of heteronuclear triple resonance NMR spectroscopy. The ^{1}H, ^{15}N and ^{13}C assignment obtained for S100A5m are listed in Table 3. Backbone assignments for amide resonances were obtained for 82 residues out of 92 residues (89.1%). Unassigned residues did not show backbone amide resonance in the 2D ^{1}H–^{15}N HSQC spectrum. The assignments are annotated with residue types and numbers in Figure 8.

NMR Backbone assignment of RAGEv

Backbone assignment of RAGEv consisting of 103 amino acids with molecular weight of 11.4 kDa has also carried out using a suite of heteronuclear triple resonance

NMR spectroscopy. The ^{1}H, ^{15}N and ^{13}C assignment obtained for RAGEv are listed in Table 4. Backbone assignments for amide resonances were obtained for 82 residues out of 103 residues (79.6%). Unassigned residues did not show backbone amide resonance in the 2D ^{1}H–^{15}N HSQC spectrum or due to lacks of sequential connectivity from the NMR backbone assignment experiment. The assignments are annotated with residue types and numbers in Figure 9.

Figure 7. Superimposed ^{1}H-^{15}N HSQC spectra of S100A5 and S100A5m

^{1}H-^{15}N HSQC spectra of wild-type of S100A5 (black) and mutant of S100A5 (red) in 20 mM Tris-HCl, pH 7.4 100 mM NaCl, 5 mM BME indicates that S100A5m is well folded and has similar structure with wild-type of S100A5.

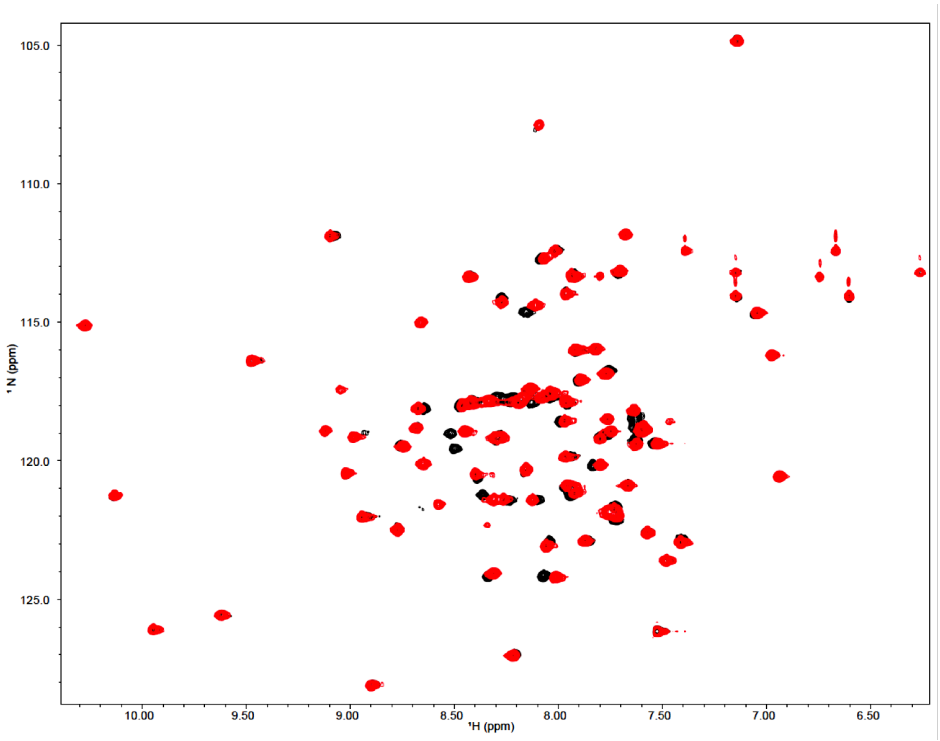

105.0
110.0
115.0
120.0
125.0
N (ppm)
10.00
9.50
9.00
8.50
8.00
7.50
7.00
6.50
H (ppm)

Figure 8. ^{1}H-^{15}N HSQC spectrum of S100A5m

1 mM ^{13}C,^{15}N-labeled S100A5m in 20 mM Tris-HCl, pH 7.4, 100 mM NaCl, at 25°C. The backbone amide are annotated with the residue types and numbers.

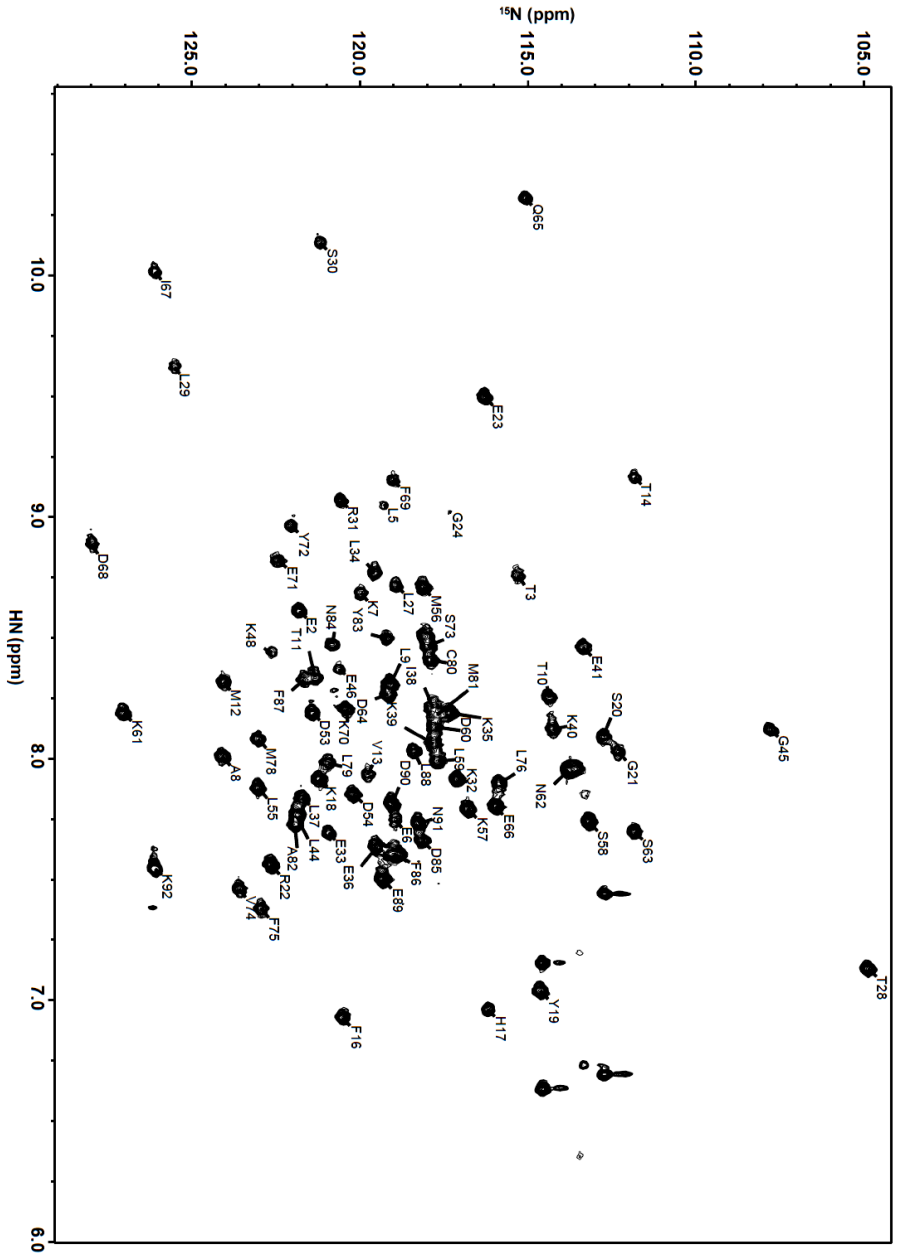

^{15}N (ppm)
105.0
110.0
115.0
120.0
125.0
HN (ppm)
6.0
7.0
8.0
9.0
10.0
Q65
S30
I67
L29
E23
T14
F69
L5
R31
G24
T3
D68
Y72
L34
E71
K7
Y83
L27
M56
S73
C80
M81
N84
E2
T11
K48
L9
I38
E46
K39
T10
E41
S20
K40
G21
G45
M12
F87
D64
K70
D90
K35
N62
K61
D53
L79
V13
K32
L76
E66
K57
A8
M78
K18
D54
N91
S58
S63
L55
L37
L44
E33
E6
D85
E36
F86
A82
R22
E89
K92
Y74
F75
T28
Y19
H17
F16

Figure 9. ^{1}H-^{15}N HSQC spectrum of RAGEv

0.7 mM ^{13}C,^{15}N-labeled RAGEv in 20 mM Tris-HCl, pH 7.4, 100 mM NaCl, at 25°C. The backbone amide are annotated with the residue types and numbers.

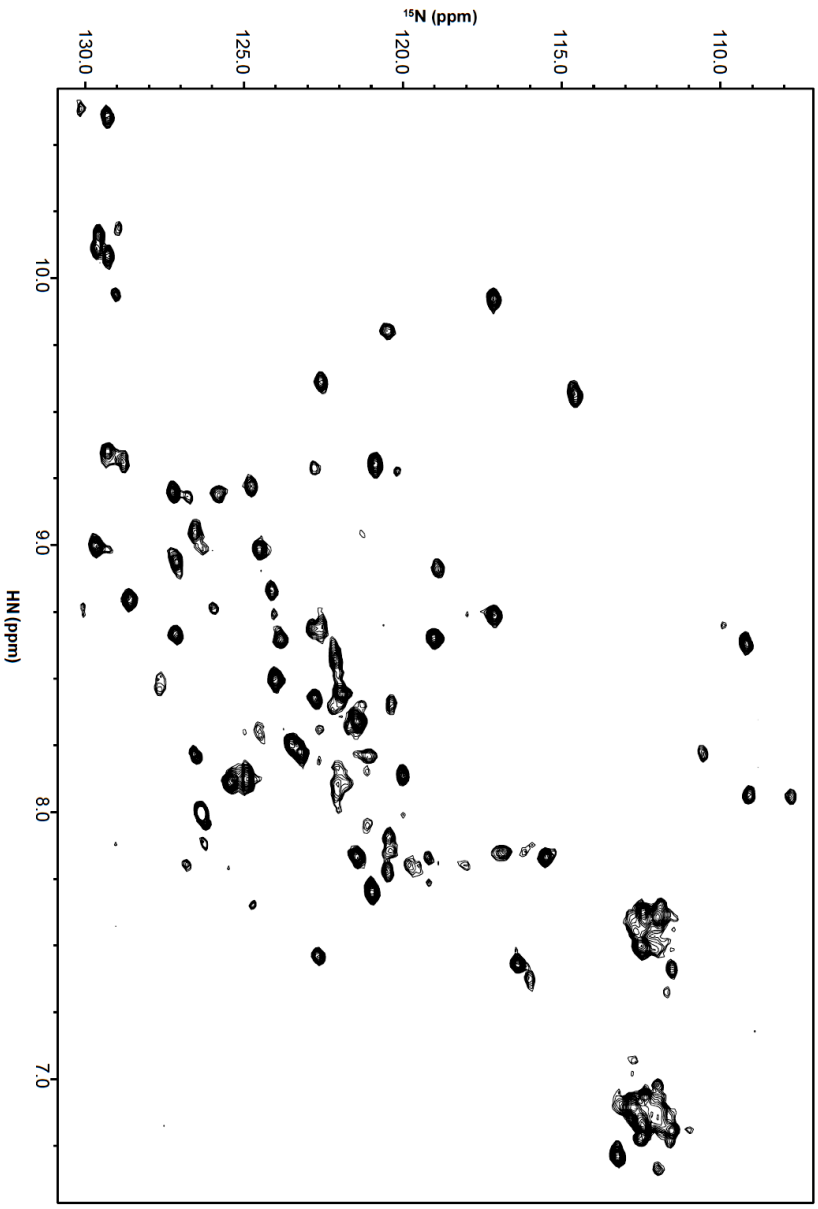

15N (ppm)
130.0
125.0
120.0
115.0
110.0
HN (ppm)
10.0
9.0
8.0
7.0

Table 3. Backbone ^{1}H, ^{15}N, $^{13}C_{\alpha}$, $^{13}C_{\beta}$, and ^{13}CO chemical shifts of S100A5m. (unit ppm)

Residue	^{1}H	^{15}N	$^{13}C_{\alpha}$	$^{13}C_{\beta}$	^{13}CO
M1					
E2	8.61	121.80	56.96	31.51	177.06
T3	8.76	115.30	57.06	60.54	173.20
P4					
L5	9.05	119.28	58.16	42.23	178.27
E6	7.75	118.94	60.26	29.55	179.87
K7	8.69	119.96	59.77	32.19	179.42
A8	8.01	124.10	55.39	18.68	179.21
L9	8.27	119.16	58.14	41.38	179.47
T10	8.25	114.36	67.28	68.36	
T11	8.34	121.33	67.03	68.10	
M12	8.32	124.06	59.95	33.16	178.56
V13	7.93	119.77	66.98	32.90	177.78
T14	9.16	111.83	66.35	68.65	
T15	8.18	117.70			
F16	6.93	120.51	62.64	39.19	176.48
H17	6.96	116.18	57.92	31.58	177.50
K18	7.92	121.22	58.81	32.62	177.67
Y19	7.04	114.63	59.89	40.57	175.76
S20	8.09	112.76	61.49		
G21	8.03	112.30	45.63		173.96
R22	7.56	122.64	59.10	30.55	177.15
E23	9.49	116.28	54.32	33.07	175.87
G24					
S25					
K26					
L27	8.72	118.91	54.05	42.95	175.66
T28	7.13	104.87	59.08	72.91	173.60

Residue	^{1}H	^{15}N	$^{13}C_{\alpha}$	$^{13}C_{\beta}$	^{13}CO
L29	9.62	125.52	52.53	43.44	176.32
S30	10.14	121.18	56.31	65.58	174.29
R31	9.07	120.57	60.50	30.36	177.79
K32	7.91	117.10	60.04	33.10	179.70
E33	7.69	120.92	59.46	30.65	178.00
L34	8.77	119.55	57.91	41.40	177.71
K35	8.19	117.33	60.49	32.64	178.00
E36	7.63	119.50	59.41	29.28	176.55
L37	7.77	121.87	59.32	42.25	178.99
I38	8.21	117.82	65.97	38.44	177.68
K39	8.06	117.80			178.64
K40	8.12	114.25	57.03	33.78	179.21
E41	8.46	113.36	55.79	30.48	176.76
L42					
S43					
L44	7.84	121.73	55.22	42.05	177.70
G45	8.12	107.77	45.38		173.71
E46	8.37	120.61	56.71	29.69	176.54
M47	8.25	123.96			175.96
K48	8.44	122.64	56.30	33.37	177.85
E49					
S50					
S51	7.40	118.71	60.75	62.67	
I52					
D53	8.19	121.39	58.11	40.02	179.03
D54	7.85	120.19	57.56	40.31	178.68
L55	7.88	123.02	57.82	41.98	178.87
M56	8.71	118.12	58.12	30.78	177.93
K57	7.79	116.78	59.28	32.30	178.99
S58	7.74	113.18	60.95	63.46	

Residue	^{1}H	^{15}N	$^{13}C_{\alpha}$	$^{13}C_{\beta}$	^{13}CO
L59	7.99	117.68	55.40	44.72	178.43
D60	8.13	117.77	53.41	39.09	177.45
K61	8.19	127.04	57.91	32.59	177.52
N62	7.95	113.80	51.68	37.03	174.59
S63	7.70	111.82	59.34	61.66	
D64	8.31	119.11	52.55	40.61	176.70
Q65	10.32	115.08	57.62	25.91	173.44
E66	7.80	115.96	53.93	33.29	174.92
I67	10.02	126.11	57.20	36.58	176.41
D68	8.89	127.98	52.61	41.45	175.24
F69	9.16	119.02	63.48	39.31	177.68
K70	8.20	120.40	60.46	31.74	180.25
E71	8.82	122.44	59.13	30.20	179.90
Y72	8.97	122.05	60.24	38.80	176.73
S73	8.51	118.04			
V74	7.46	123.60	66.42	31.40	178.82
F75	7.38	122.94	58.89	38.07	176.99
L76	7.90	115.84	57.98	40.51	179.81
T77					
M78	8.08	123.03	59.68	32.61	179.95
L79	7.98	120.95	57.85	41.84	179.48
C80	8.47	117.99	65.09	26.76	176.68
M81	8.41	117.90	58.99	31.81	178.11
A82	7.73	121.93	55.02	17.83	180.42
Y83	8.50	119.20	61.34	37.42	178.60
N84	8.47	120.83	57.30	38.45	176.08
D85	7.66	118.14			179.13
F86	7.61	118.99	60.03	38.54	177.54
F87	8.33	121.66	60.74	39.35	
L88	8.03	118.37	56.40	42.18	178.63

Residue	^{1}H	^{15}N	$^{13}C_{\alpha}$	$^{13}C_{\beta}$	^{13}CO
E89	7.50	119.32	57.88	29.76	
D90	7.81	119.05	54.89	41.43	176.03
N91	7.73	118.26	53.29	39.04	183.50
K92	7.54	126.09	57.99	46.33	

Table 4. Backbone ^{1}H, ^{15}N, $^{13}C_{\alpha}$, and $^{13}C_{\beta}$ chemical shifts of RAGEv.

(unit ppm)

Residue	HN	N	C_{α}	C_{β}
Q24	8.44	121.93	56.70	33.45
N25	8.99	125.96		
I26	9.20	127.22	58.58	39.92
T27	8.69	122.83	61.00	69.81
A28	9.00	129.65	50.22	23.11
R29	8.34	121.42	54.92	31.35
I30	8.23	123.24	62.96	38.36
G31	9.57	114.58	45.00	
E32	7.70	120.97	54.17	29.85
P33				
L34	8.57	122.12	54.34	46.90
V35	8.12	124.91	60.57	33.59
L36	8.94	127.12	52.25	44.34
K37	8.45	121.96	56.73	33.61
C38	9.00	127.35	56.46	44.03
K39	8.21	126.50	58.47	31.02
G40	8.40	109.48	44.90	
A41	7.46	122.66	50.19	18.21
P42				
K43				
K44				
P45				
P46				
Q47				
R48	9.13	125.57	55.29	29.53
L49	7.83	121.43	54.09	45.15
E50	8.41	120.36	56.07	34.64
W51	9.19	125.81	56.72	32.52
K52	9.80	120.47	54.94	36.92
L53	8.43	119.48	54.28	44.94
N54	8.17	117.16	52.26	
T55	8.09	109.20	59.03	72.60
G56	8.00	118.39	47.45	
R57	7.91	109.43	57.08	32.38
T58	7.91	109.43	61.15	70.22

Residue	HN	N	C_α	C_β
E59				
A60				
W61	9.43	126.53	58.27	29.46
K62	9.43	126.53	54.05	33.02
V63	9.28	120.18	60.24	33.71
L64				
S65	8.73	118.40	55.81	64.41
P66				
Q67	8.28	117.19	56.66	28.73
G68	8.19	108.85	45.59	
G69	9.16	110.93	45.11	
G70	8.51	109.92	45.00	
P71				
W72	8.65	117.82	59.36	28.06
D73	7.90	120.43	57.05	40.62
S74	7.77	111.61	59.30	63.75
V75	7.80	119.80	63.28	33.86
A76	8.24	121.19	51.68	20.71
R77	8.94	118.95	54.00	34.22
V78	8.65	123.85	62.29	31.91
L79	9.35	129.29	54.00	39.51
P80				
N81	7.34	110.23	52.01	37.20
G82	8.63	109.20	45.63	
S83	7.83	115.51	59.34	63.83
L84	8.83	124.12	53.62	41.75
F85				
L86	8.77	130.07	52.01	42.03
P87				
A88	7.38	116.01	51.77	19.34
V89	8.12	125.39	64.19	32.30
G90	9.93	117.15	43.49	
I91	8.65	119.02	63.94	38.27
Q92	8.74	117.12	57.41	28.39
D93	7.85	116.85	55.98	41.92
E94	6.71	113.23	57.60	30.66
G95	8.22	110.55	45.26	
I96	8.02	121.12	59.25	40.32
F97	9.22	124.76	56.62	43.42

Residue	HN	N	C_α	C_β
R98				
C99				
Q100	8.46	127.66	54.99	31.62
A101	8.66	127.12	50.52	22.64
M102	8.14	120.00	54.18	33.90
N103	8.43	122.75	51.53	39.25
R104	8.76	124.02	58.73	29.68
N105	8.00	115.86	53.11	38.80
G106	8.05	107.80	45.82	
K107	7.78	120.49	55.04	33.32
E108	8.76	125.95	55.65	32.01
T109	9.04	121.42	62.29	70.37
K110	8.87	127.59	55.57	35.03
S111				
N112				
Y113	9.61	122.57	56.93	42.30
R114	8.98	124.52	54.91	31.51
V115	9.31	128.80	62.35	33.67
R116	8.80	128.60	54.34	33.19
V117	9.31	120.87	60.59	33.48
Y118	8.50	124.03	56.99	40.59
Q119	8.40	122.27	54.96	30.30
I120	8.25	123.49	59.08	38.65
P121				
G122	8.36	108.80		
K123	8.10	122.06	54.17	32.81
P124				
E125	8.00	126.33	58.06	31.47

4.3. Calcium-dependent binding and stoichiometry between S100A5 and RAGEv

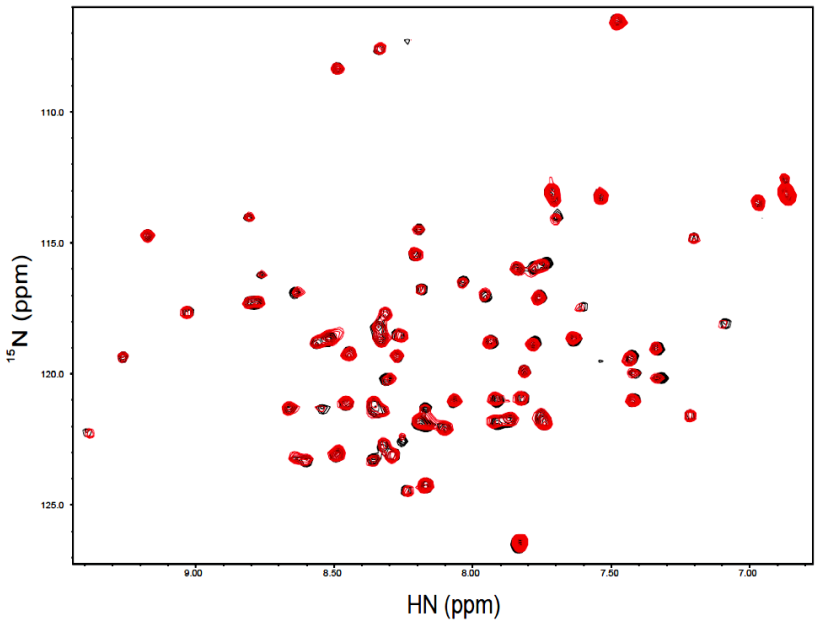
Because calcium induces the large conformational change in S100A5, I monitored ^{1}H-^{15}N HSQC spectra from the complex between S100A5 and RAGEv in the absence and presence of calcium, respectively. Only calcium-bound S100A5m with RAGEv exhibited most backbone amide resonances broaden out, which indicates interaction between S100A5 and RAGE is dependent on calcium (Fig. 10). I performed experiments for characterization of the interaction in the presence of calcium hereafter.

To figure out the stoichiometry of the complex of S100A5m and RAGEv, I measured the complex molecular mass using size exclusion chromatography (SEC). The calculated molecular mass of S100A5m and RAGEv were 10.801 and 11.554 Da, respectively. The S100A5m dimer eluted as a single peak at 12.7 ml with a molecular mass of ~21 kDa, and RAGEv eluted as a monomer at 13.7 ml with a molecular mass of 14.1 kDa (Fig. 11 A and B). The 1:1 mixture of the S100A5m and RAGEv eluted as two peaks at 11.6 ml and 13.7 ml with molecular mass 32.6 kDa and 14.1 kD, respectively. The molecular mass corresponds to a heterotrimer of $(S100A5m)_2$:RAGEv and free excess RAGEv (Fig. 11C). I further confirmed the absolute molecular mass of the complex using the multi-angle light scattering (MALS) and refractive index data. The molecular mass of the S100A5 dimer, RAGEv and the complex was determined as 20.9 ± 0.4, 13.3 ± 0.4, 30.9 ± 0.4 kDa, respectively, which was consistent with the heterotrimeric complex between the S100A5m dimer and RAGEv (Fig. 11 A-C). The ratio of RAGEv to S100A5 was increased, but elution volume and molar mass did not change, indicating that the

complex is formed as 2:1 stoichiometry of the $(S100A5m)_2$:RAGEv heterotrimer. I then examined the equilibrium binding thermodynamics of S100A5m and RAGEv using calorimetry. The binding between S100A5m and RAGEv was largely endothermic reaction, with the K_D value of 5.9 ± 0.3 μM, ΔH of 7.5 ± 0.1 kcal/mol and $-T\Delta S$ of -14.7 ± 0.1 kcal/mol, respectively. The large endothermic ΔH value indicates that the association between S100A5m and RAGEv involves mainly hydrophobic interfaces. The stoichiometry in the titration of RAGEv into S100A5 dimer was measured as 0.82 close to 1, corroborating the heterotrimer formation between RAGEv and the S100A5 dimer.

Figure 10. ^{1}H-^{15}N HSQC spectrum of S100A5m with RAGEv in the (top) absence and (bottom) presence of Ca^{2+}.

Top panel: 0.2 mM ^{15}N-labeled S100A5m alone (black) with 0.4 mM unlabeled RAGEv (red) in 20 mM Tris-HCl, pH 7.4, 100 mM NaCl, 5 mM EDTA. Bottom panel: same concentration of protein samples in 20 mM Tris-HCl, pH 7.4, 100 mM NaCl, 5 mM $CaCl_2$. In the presence of calcium, most backbone amide resonances were broadened out.


110.0
115.0
120.0
125.0
^{15}N (ppm)
9.00
8.50
8.00
7.50
7.00
HN (ppm)

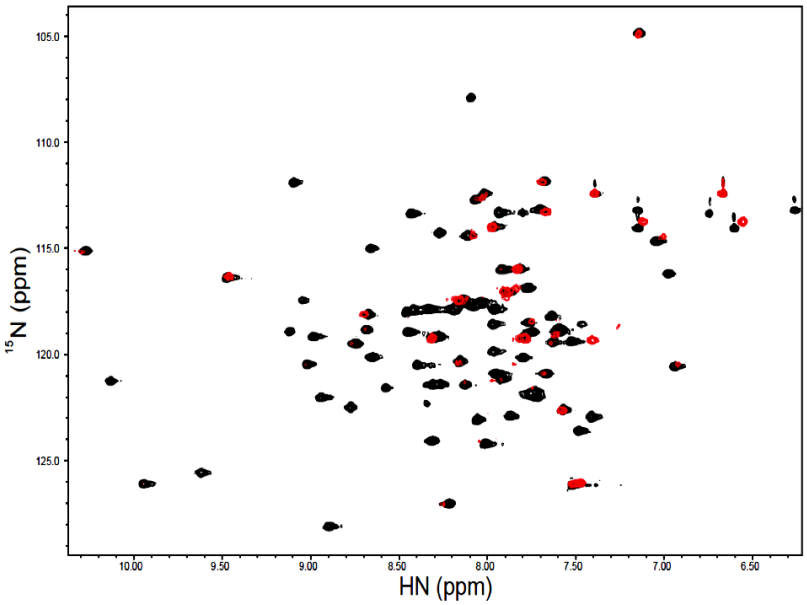

105.0
110.0
115.0
120.0
125.0
^{15}N (ppm)
10.00
9.50
9.00
8.50
8.00
7.50
7.00
6.50
HN (ppm)

Figure 11. Size exclusion chromatograms of S100A5, RAGEv and S100A5-RAGEv complex.

(A) S100A5m, (B) RAGEv, and (C) the S100A5m-RAGEv complex in 20 mM Tris-HCl, pH 7.4, 100 mM NaCl, 10 mM $CaCl_2$ are eluted using a Superdex 75 10/300 GL column (GE healthcare) at 25°C. The proteins were confirmed by SDS-PAGE in the bottom panel.

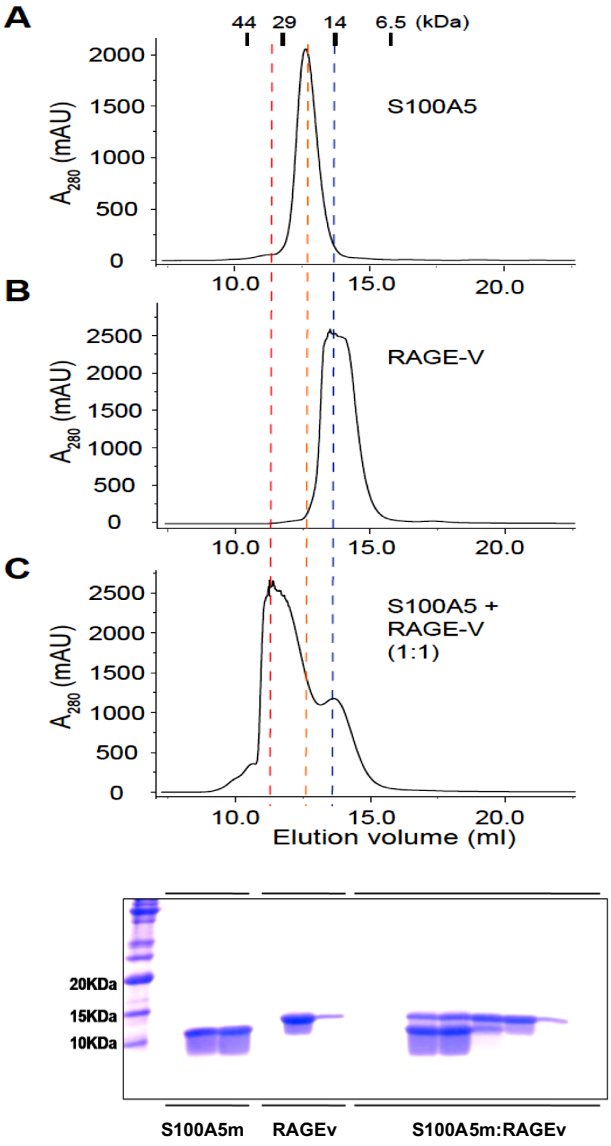

A
44
29
14
6.5 (kDa)
2000
1500
1000
500
0
A280 (mAU)
S100A5
10.0
15.0
20.0
B
2500
2000
1500
1000
500
0
A280 (mAU)
RAGE-V
10.0
15.0
20.0
C
2500
2000
1500
1000
500
0
A280 (mAU)
S100A5 +
RAGE-V
(1:1)
10.0
15.0
20.0
Elution volume (ml)
20KDa
15KDa
10KDa
S100A5m
RAGEv
S100A5m:RAGEv

Figure 12. Elution profiles of S100A5, RAGEv and S100A5-RAGEv complex from MALS.

The molecular masses of (A) S100A5m, (B) RAGEv, and (C) S100A5m-RAGEv were obtained from the light scattering and refractive index measurements correspond to free S100A5m dimer, RAGEv, and the $(S100A5m)_2$:RAGEv heterotrimer complex in the chromatic separation from WTC-030S5 column (Wyatt).

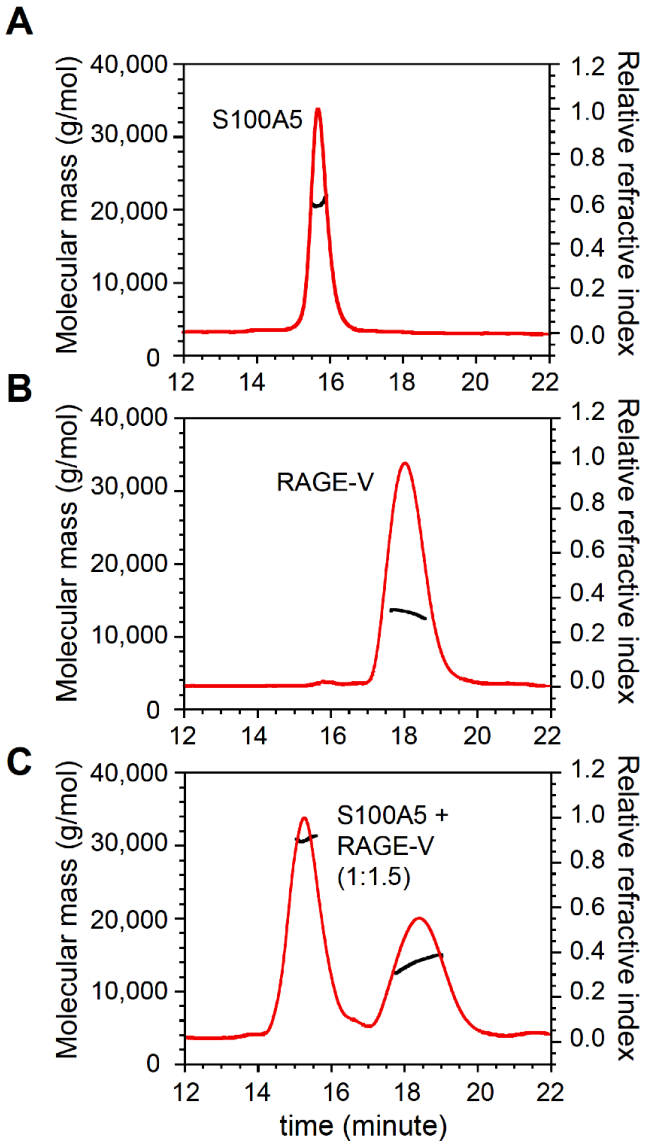

A
B
C
S100A5
RAGE-V
S100A5 +
RAGE-V
(1:1.5)
Molecular mass (g/mol)
Relative refractive index
time (minute)
40,000
30,000
20,000
10,000
0
1.2
1.0
0.8
0.6
0.4
0.2
0.0
12
14
16
18
20
22

Figure 13. Thermodynamics between S100A5 homodimer and RAGEv derived from ITC.

Raw ITC data (top panels) and integrated heats of injection (bottom panels) for the titration of S100A5 and RAGEv at 25°C. The squares in the bottom panels are the experimental data, and the lines represent the least-squares best-fit curves derived for a one-site binding model. The thermodynamic parameters are given in Table 5.

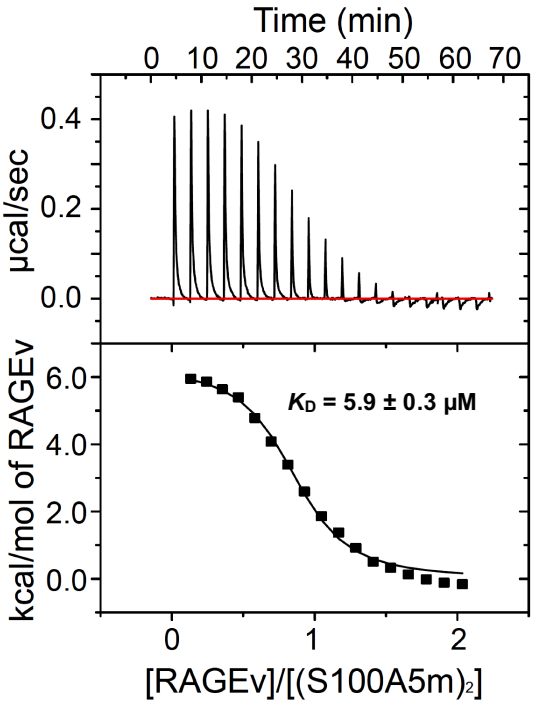

Time (min)
0 10 20 30 40 50 60 70
µcal/sec
0.4
0.2
0.0
kcal/mol of RAGEv
6.0
4.0
2.0
0.0
K_D = 5.9 ± 0.3 µM
0
1
2
[RAGEv]/[(S100A5m)$_2$]

Table 5. Equilibrium dissociation constants (K_D), free energy changes (ΔG), enthalpy changes (ΔH), and entropy changes ($T\Delta S$) for the interaction with RAGEv of S100A5m

	K_D (μM)	ΔG (kcal/mol)	ΔH (kcal/mol)	$T\Delta S$ (kcal/mol)
S100A5m:RAGEv	5.9 ± 0.3	−7.2 ± 0.1	7.5 ± 0.1	14.7 ± 0.1

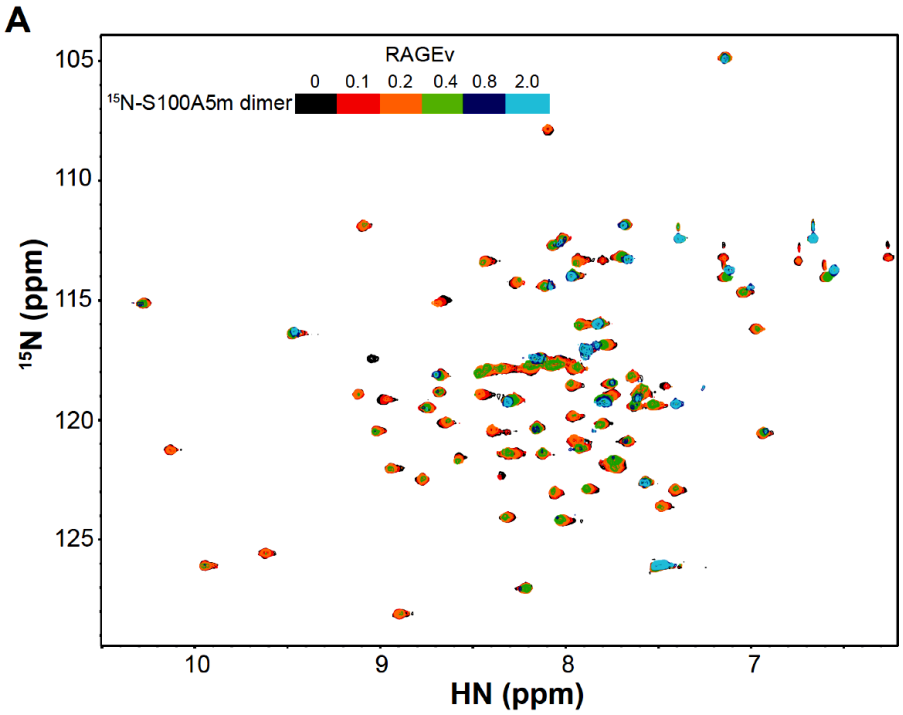
4. 4. Characterization of the complex of S100A5 and RAGEv

I performed NMR titration between S100A5m and RAGEv to map the binding interface. ^{15}N-S100A5m exhibited a severe line broadening in the middle of the titration with RAGEv, such that most backbone amide resonances were broaden out beyond detection with little chemical shift changes (Fig. 14A). Similarly, ^{15}N RAGEv showed a significant line broadening during the titration with S100A5m, and most backbone amide resonances largely broadened out upon complex formation (Fig 15A). NMR titration indicates a prevalence of an intermediate exchange in the chemical shift time scale, hence a dynamic nature of the complex. I carried out the titration between S100A5m and RAGEv at varying temperatures between 5℃ and 45℃, but line broadening was persistent. The intensity ratio of backbone amide resonances between free and complex state showed that most residues exhibited line broadening and intensity decrease upon complex formation (Fig. 14B and 15B). I selected residues with large intensity decrease to identify the binding interface. The residues with intensity ratio below 0.3 and 0.5 were selected for S100A5m and RAGEv, respectively. Thr3, Leu5, Glu6, Lys7, Thr14, Leu42, Met78, Leu79, Tyr83, Asn84 and Phe86 of S100A5 showed the largest decrease from the titration. The residues were mainly hydrophobic and located in helix I and IV of S100A5m around the interface for dimerization. Ala28, Ile30, Thr55, Gly56, Thr58, Val75, Arg77, Leu79, Gln100 and Lys110 of RAGEv showed the largest intensity decrease. The selected residues were shown on the three dimensional structures of S100A5 and RAGEv to map the binding interfaces (Fig. 16 C and D). The identified binding interfaces suggest that S100A5 employ the periphery of the dimer interface around helix I and IV to interact with RAGEv. The binding mode is

consistent with the heterotrimeric formation between RAGEv and an S100A5 dimer. The interaction employs hydrophobic surfaces at the dimer interface of S100A5, which is exposed by conformational change upon Ca^{2+}-binding. Among others, side chain of Met78 and Tyr83 are exposed by Ca^{2+}-binding, and can contribute significantly to the hydrophobic interaction with RAGEv.

Figure 14. Superimposed ^{1}H-^{15}N HSQC spectra and intensity ratio histogram from NMR titration of S100A5m with RAGEv.

(A) 0.4 mM ^{15}N-labeled S100A5m was stoichiometrically titrated with unlabeled RAGEv according to the molar ratio that is color-coded as the bar graph in the spectra. (B) The intensity ratio of backbone amide resonances in ^{1}H-^{15}N HSQC spectra between 50% bound complex and the free protein as a function of residue number upon complex formation between S100A5m and RAGEv


A
RAGEv
0 0.1 0.2 0.4 0.8 2.0
15N-S100A5m dimer
105
110
115
120
125
15N (ppm)
10
9
8
7
HN (ppm)

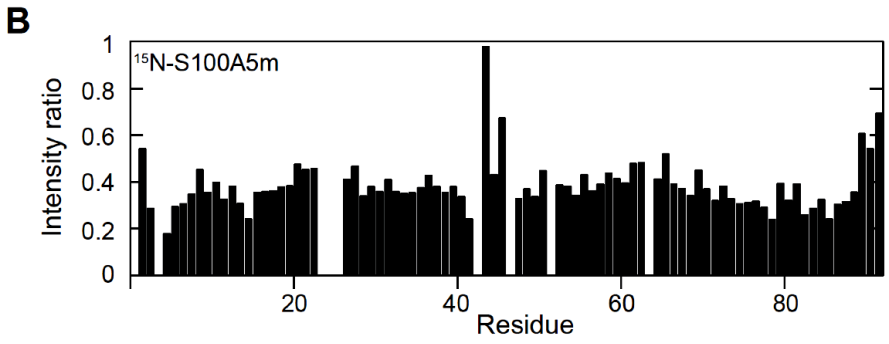
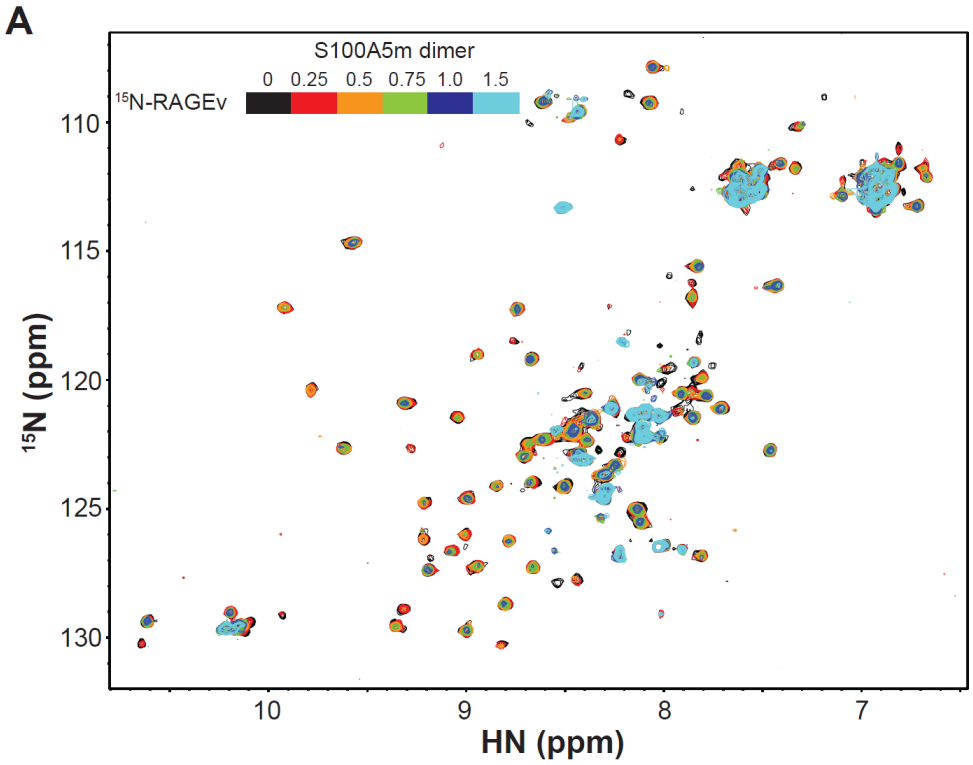

B
15N-S100A5m
1
0.8
0.6
0.4
0.2
0
Intensity ratio
20
40
60
80
Residue

Figure 15. Superimposed ^{1}H-^{15}N HSQC spectra and intensity ratio histogram from NMR titration of RAGEv with S100A5m.

(A) 0.4 mM ^{15}N-labeled RAGEv was stoichiometrically titrated with unlabeled S100A5m according to the molar ratio that is color-coded as the bar graph in the spectra. (B) The intensity ratio of backbone amide resonances in ^{1}H-^{15}N HSQC spectra between 50% bound complex and the free protein as a function of residue number upon complex formation between S100A5m and RAGEv


A
S100A5m dimer
0 0.25 0.5 0.75 1.0 1.5
15N-RAGEv
110
115
120
125
130
15N (ppm)
10
9
8
7
HN (ppm)

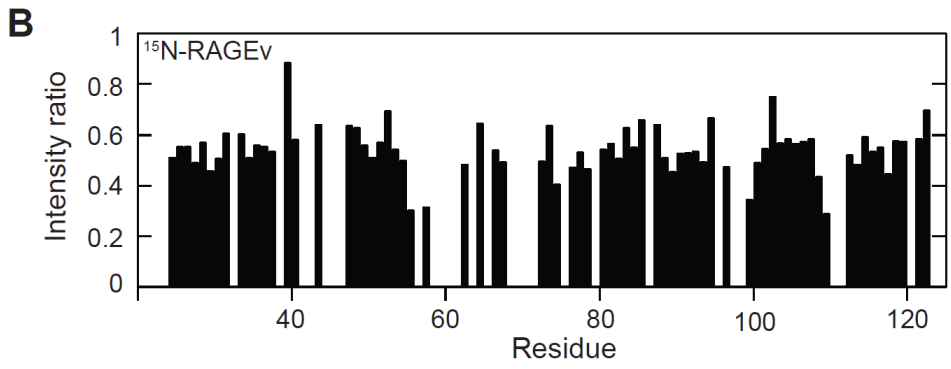
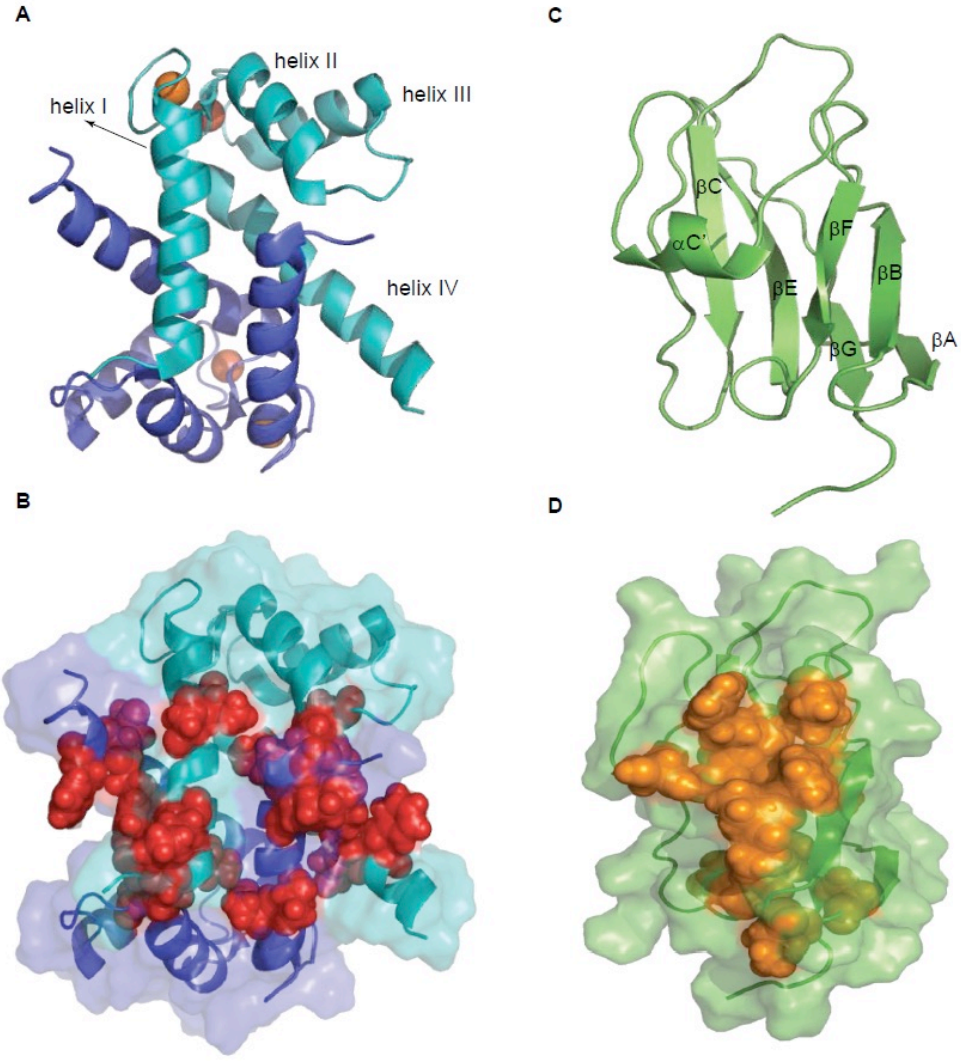

B
15N-RAGEv
1
0.8
0.6
0.4
0.2
0
Intensity ratio
40
60
80
100
120
Residue

Figure 16. Binding interfaces between S100A5 and RAGEv.

(A and C) S100A5 dimer and (B and D) RAGEv are shown as ribbon diagram and transparent surface model (PDB code 2kax for S100A5 2e5e for RAGEv). Residues with large intensity decrease from NMR titration are shown on (C) S100A5 as red sphere and (D) RAGEv as orange spheres.


A
helix I
helix II
helix III
helix IV
C
βC'
αC'
βF
βB
βE
βG
βA
B
D

5. Discussions

I quantitatively determined Ca^{2+}-affinity of each EF-hand, such that equilibrium dissociation constant for Ca^{2+}-binding of the C-terminal EF-hand is a ~3 times lower than that of the N-terminal EF-hand, indicating higher Ca^{2+} binding affinity of C-terminal EF-hand. Previous study using flow dialysis showed that Ca^{2+} ions bind to S100A5 in a highly cooperative manner[20]. Intrinsic Ca^{2+}-affinity of the N-terminal EF-hand in the apo-state could be much weaker, considering the cooperative nature of Ca^{2+}-binding of S100A5. It has been suggested that water molecules around Ca^{2+}-binding site and short β-sheet connecting the C-terminal and N-terminal EF-hands contribute to the Ca^{2+}-binding cooperativity in the crystal structure of S100A6[38]. Since S100A5 has a sequence and structural homology to S100A6, the cooperativity can be comparable. Given the binding enthalpy and entropy of Ca^{2+}-binding at the C-terminal EF-hand in S100A5, enthalpic and entropic contribution of C-terminal EF-hands might be a sum of enthalpic and entropic cost of major conformational rearrangements including reorientation of Helix III and elongation of Helix IV, and an exposure of hydrophobic patches and displacement of water molecules within the Ca^{2+}-coordination by Thr28, Asp60, Asn62, Asp64, and Glu66 in S100A5. On the other hand, Ca^{2+}-binding at the N-terminal EF-hand might be a sum of a minor conformational change in Helix I and a water molecule displacement within the Ca^{2+}-coordination by Ser20, Arg22, Glu23 and Leu29. I speculate that Thr28 and Leu29 within a short β-sheet between C-terminal and N-terminal EF-hand contributes to the cooperativity between C-terminal and N-terminal EF-hand, thereby increasing

Ca^{2+}-affinity of the N-terminal EF-hand from intrinsic submilimolar to submicromolar affinity.

The heterotrimer formation of S100A5 family proteins with RAGE is less known than the heterotetramer formation. Previously S100A6 and S100P was shown to form asymmetric heterotetramer with RAGEv, such that individual S100 dimers associated with two RAGEv[33,34]. The K_D values of S100A6 and S100P for RAGEv binding were measured as 3-6 μM, which comparable to that of S100A5. NMR titration of S100A5 or S100P with RAGEv showed line broadening that was limited to the binding interfaces. Recent study have suggested that modeled complex between S100A5 and RAGEv as a heterotetramer, which is not consistent with this study[39]. It remains to be clarified whether S100A5 interacts RAGEv with formation of multiple complex states at different buffer conditions.

The intracellular signaling in the S100A5-RAGE seems to be induced by heterotrimer formation rather than heterotetramer. I speculate that S100A5 dimer presumably replace a ligand which previously binds RAGE by forming the heterotrimeric complex, thus interrupting RAGE dimerization, which alters the intracellular signaling. Although the binding interface for S100A5-RAGEv can be mapped onto the surface of S100A5 and RAGEv, respectively, the binding mechanism remains unclear. I speculate two possible ways. One way is that the binding interface on the surface of S100A5 might be entirely covered by RAGE, thus forming a heterotrimer complex as stoichiometrically 1:1. The other is that the interface for homodimerization of S100A5 might be distorted with allosterically when a RAGEv molecule binds to S100A5, thus disturbing additional RAGEv binding. It is inferred that the dynamic nature of the heterotrimer complex is much

larger than the heterotetramer complex. Give that the S100 family proteins form a dimer with similar backbone folds, it is puzzling what determines the stoichiometry of the complex. Most of S100 proteins associated with RAGE only as holo states, and I speculate that the distribution of hydrophobic and charged residues altered upon Ca^{2+}-binding would be an important factor for the assembly. Distinct binding modes of RAGE against various S100 family proteins may help to discriminate structurally homologous ligand proteins to control proper RAGE signaling.

6. Conclusion

Since calcium ion is an important 2nd messenger as a signal in the cell, it is important to understand features of calcium-binding and signaling, such as transduction path, affinity, and cooperativity. In this study, calcium-binding of S100A5 and the interaction between calcium binding protein S100A5 and RAGE were characterized. S100A5 exhibits calcium-binding in the sequential manner using each C-terminal and N-terminal EF-hand with K_D of 1.3 μM and 3.5 μM, respectively. In addition, calcium-binding protein S100A5 interacts with the V domain of RAGE (RAGEv) to form a heterotrimeric complex with K_D of 5.9 μM in a strict calcium-dependent manner, which is a distinct stoichiometry from the S100 family proteins. Chemical shift perturbation data from NMR titration experiments indicates that S100A5 employs the periphery of the dimer interface to interact with RAGEv. This study could broaden understanding nature of calcium-binding molecule to modulate a calcium-associated signaling and different distinct binding mode and stoichiometry of S100 family protein against RAGE could be important to modulate diverse RAGE signaling.

CHAPTER II

Solution structure and dynamics of anti-CRISPR AcrIIA4, Cas9 inhibitor

1. Abstract

The bacterial CRISPR-Cas system provides adaptive immunity against invading phages and foreign plasmids. Cas9, an RNA-guided endonuclease, cleaves target DNA substrates on the specific site and constitutes a well-established platform for genome editing. Recently, anti-CRISPR (Acr) proteins that inhibit Cas9 have been discovered, promising a useful off-switch for Cas9 to avoid undesirable off-target effects. Here, I report the solution structure and dynamics of *Listeria monocytogenes* AcrIIA4 that inhibits *Streptococcus pyogenes* Cas9 (*Spy*Cas9). AcrIIA4 forms a compact monomeric αβββαα fold comprising three antiparallel β strands flanked by three α-helices and a short 3_{10}-helix. AcrIIA4 exhibits distinct backbone dynamics in fast and slow time scales at loop regions that form the interaction surface for *Spy*Cas9. In particular, the β1- β2 loop that binds to the RuvC domain of *Spy*Cas9 is highly mobile, and the β1- β2 and α2-α3 loops that bind to the C-terminal domain of *Spy*Cas9 undergoes conformational exchanges in microsecond-to-millisecond time scales. AcrIIA4 binds to apo-*Spy*Cas9 with K_D ~4.8 μM, which compares to K_D ~0.6 nM for AcrIIA4 binding to single guide RNA (sgRNA)-bound *Spy*Cas9. Since AcrIIA4 does not compete with the target DNA binding in the absence of sgRNA for the binary complex between AcrIIA4 and *Spy*Cas9 with the moderate affinity it can effectively disable the Cas9 nuclease activity by forming a tight ternary complex in the presence of sgRNA.

2. Introduction

2. 1. CRISPR-Cas9

The clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR-associated (Cas) proteins is one of the immune systems in bacteria and archea against invasion by foreign nucleic acids such as phages or plasmids[40-44]. Upon phage invasion, CRISPR-Cas system is activated by the bacterial host cell, and short segments of the invading nucleic acids are integrated into CRISPR loci. The integrated DNA sequence, known as spacers, are transcribed into CRISPR RNAs (crRNA) that guide a single protein effector or a multi-protein effector complex to target and cleave the complementary sequences in foreign genes upon next invasion. CRISPR-Cas systems are grouped into two classes (class 1 and 2), six types and 21 subtypes according to different organizations of signature *cas* genes[45,46]. *Streptococcus pyogenes* Cas9 (*Spy*Cas9) is classified to the subtype II-A of the Class 2 CRISPR-Cas system. *Spy*Cas9 has been extensively investigated and widely applied to genome editing due to its simplicity in generating double-strand DNA breaks at desired sites targeted by single-guide RNA (sgRNA)[46]. General CRISPR-Cas9 mechanism is depicted (Fig. 1)

2. 2. Anti-CRISPR proteins

Bacteria and phages have long been engaged in an arms race, developing a variety

of arsenal to compete for invasion and defense[47]. The arms races between run with the Red Queen hypothesis in that bacteria and phages must continually evolve their mechanism to survive. To date, five CRISPR-Cas9 inhibition mechanisms and 21 unique families of anti-CRISPR (Acr) proteins against type I and type II CRISPR-Cas systems has been described[48]. In spite of their small size, there are no common features such as sequence and structural homology between them. Among them, *Listeria monoycytogenes* prophage encodes AcrIIA2 and AcrIIA4 and these Acr proteins have shown cross-strain inhibition against *Spy*Cas9, highlighting their potential as an off-switch for Cas9 for temporal and spatial control of genome editing[49]. Recent structures of the AcrIIA4-*Spy*Cas9-sgRNA complex obtained by X-ray crystallography and electron microscopy show that *Spy*Cas9 binds to AcrIIA4 via the protospacer adjacent motif (PAM) interaction site and the RuvC domain[50-52].

Here, I investigated the solution structure and dynamics of AcrIIA4 using NMR spectroscopy. Based on backbone dynamics from ^{15}N NMR relaxation data, I demonstrates that the binding interfaces for *Spy*Cas9 exhibit fast internal motions and slow conformational exchanges in variable timescales. Further, I show that AcrIIA4 binds to *Spy*Cas9 in the absence and the presence of sgRNA both with equilibrium dissociation constants (K_D) of ~4.8 μM and 0.6 nM, respectively. AcrIIA4-*Spy*Cas9 complex can associate with guide RNA to form a tight ternary complex, leading to the loss of Cas9 nuclease activity. According to comparison for binding interfaces between AcrIIA4 and *Spy*Cas9 in the absence and presence of sgRNA, I suggest a sequential binding model for AcrIIA4-*Spy*Cas9 binary complex to AcrIIA4-*Spy*Cas9-sgRNA tight ternary complex.

Figure 1. Schematic illustration of CRISPR-Cas9 mechanism for target DNA cleavage.

The Cas9 protein remains inactivate until guide RNA binds to Cas9. Upon binding of guide RNA, Cas9 has a large conformational changes in mainly REC domain, turning to active state. Once the Cas9 proteins is activated, it searches for target DNA containing Protospacer adjacent motif (PAM) sequence and the sequence complementary to the guide RNA The Cas9 finds a potential target sequence with the appropriate PAM and RuvC and HNH nuclease domain in the Cs9 cleaves and dissociates the target DNA.

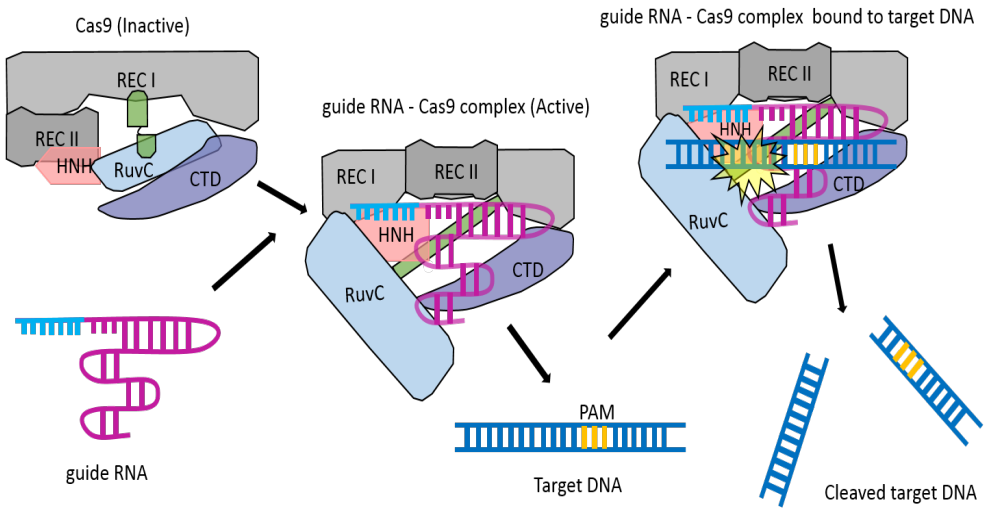
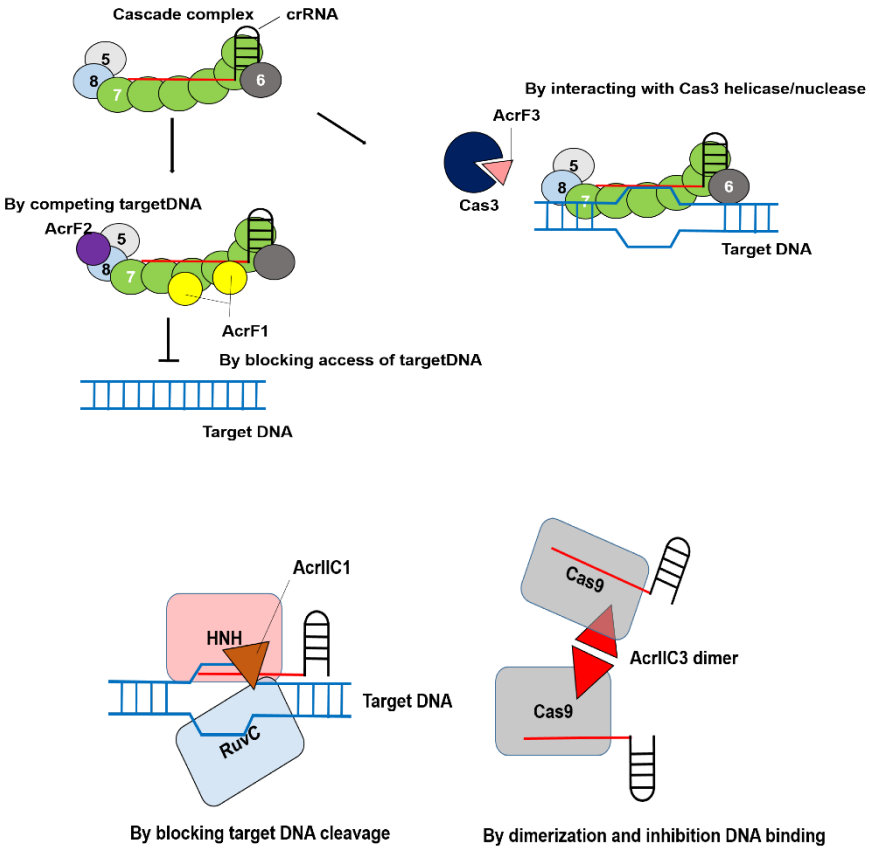

Cas9 (Inactive)
REC I
REC II
HNH
RuvC
CTD
guide RNA
guide RNA - Cas9 complex (Active)
REC I
REC II
HNH
RuvC
CTD
PAM
Target DNA
guide RNA - Cas9 complex bound to target DNA
REC I
REC II
HNH
CTD
RuvC
Cleaved target DNA

Figure 2. Strategies of CRISPR-Cas inhibition mechanism

Five CRISPR-Cas inhibition mechanism described against type I and II CRISPR-Cas system. AcrF1 binds to Cas7f protein in the Cascade complex, blocking access of the DNA target and small acidic AcrF2 protein binds to Cas8f-Cas5f in the complex and blocks DNA binding by competing with the target DNA, and AcrF3 interacts with Cas3 helicase-nuclease effector, preventing its recruitment by Cascade in type I-F system.[53-55] AcrIIC1 blocks target DNA cleavage by binding the Cas9 HNH endonuclease domain, and AcrIIC3 interferes with target DNA recognition by inducing dimerization of Cas9 in type II-C system[56]. In a mechanism similar to that of AcrF2, the AcrIIA4 targets type II-A CRISPR-Cas9 system by mimicking double-stranded DNA[50-52].


Cascade complex
crRNA
5
8
7
6
By interacting with Cas3 helicase/nuclease
AcrF3
Cas3
5
8
7
6
Target DNA
By competing targetDNA
AcrF2
5
8
7
AcrF1
By blocking access of targetDNA
Target DNA
AcrIIC1
HNH
Target DNA
RuvC
By blocking target DNA cleavage
Cas9
AcrIIC3 dimer
Cas9
By dimerization and inhibition DNA binding

3. Materials and Methods

3. 1. Cloning

The synthetic *Listeria monocytogenes* AcrIIA4 (1–87) and *Streptoccus pyogenes* Cas9 (*Spy*Cas9, 1-1368) gene was cloned into a modified pBT7-N-His vector (ATCC) and pET28b vector (Novagen) with C-terminal His_6-tag, respectively. The constructs were verified by DNA sequencing.

3. 2. Protein expression

3. 2. 1. Protein expression - Luria bertani medium

The expression of the AcrIIA4 and *Spy*Cas9 was performed using the expression vector pBT7-N-His and pET28b in *Escherichia coli* strain BL21star (DE3) (Invitrogen), respectively, using LB medium (Table 1). The cells with expression vector were cultured at 37°C to an OD_{600} of 0.6–0.8, and the protein overexpression of AcerIIA4 and *Spy*Cas9 was induced by adding IPTG to a final concentration of 1 and 0.5 mM IPTG, respectively, followed by induction for overnight at 17°C and harvested by centrifugation. The cell separated from medium was stored at -80^{o}C

3. 2. 2. Protein expression - Isotope labeling medium

[^{15}N]/[^{13}C/^{15}N]-labeled proteins are commonly used for increasing the sensitivity of NMR experiments. The overexpression vector containing AcrIIA4 gene was transformed into *Escherichia. coli* strain BL21star(DE3) (Invitrogen), and the cells

were grown in 2 ml LB medium for 2 h to inoculate to the M9 minimal media (with $^{15}NH_4Cl$ and/or $^{13}C_6$-glucose as the sole nitrogen or carbon source, respectively) (Table 1). Cells were grown at 37°C until an OD_{600} of 0.6–0.8 and induced with 1 mM IPTG for overnight grown at 17°C, harvested by centrifugation and then stored at -80^{o}C.

Table 1. The composition of media.

Medium	Composition (per Liter)
LB medium	10 g Tryptone, 5 g Yeast extract, 10 g NaCl
M9 minimal medium (^{15}N or $^{13}C/^{15}N$)	10 g K_2HPO_4, 13 g K_2HPO_4, 9 g Na_2HPO_4, 2.4 g K_2PO_4, 1 g $^{15}NH_4Cl$, 2 g U–$^{13}C6$ Glucose (or 5 g Glucose), 10 mM $MgCl_2$, 0.1 mM Thiamine, 1 x Trace element, 0.2 mM $CaCl_2$, 50 mg Carbenicillin

3. 3. Sample Preparation

3. 3. 1. AcrIIA4

The cells were harvested by centrifugation, and resuspended in buffer 20 mM Tris–HCl (pH 7.4), 200 mM NaCl, 1 mM phenylmethanesulfonyl fluoride (PMSF, Sigma, P7626), and disrupted by emulsiflex (Avestin, Canada). The lysate was centrifuged at 24,000g for 30 min at 4°C. The supernatant was loaded on a HisTrap HP column (5 ml, GE Healthcare) column equilibrated with binding buffer 20 mM Tris–HCl (pH 7.4), 200 mM NaCl and eluted with a gradient 100 ml of 500 mM Imidazole. The solution containing protein was fractionated and His_6-tag was removed using TEV protease while dialyzed against 50 mM Tris-HCl (pH 8.0), 50 mM NaCl, and 5 mM β-Mercaptoethanol (BME) for overnight. The digestion reaction was loaded onto HisTrap column. The flow-through was collected and loaded onto a Hiload Superdex 75 column (GE Healthcare) pre-equilibrated with 20 mM Tris-HCl, (pH 7.4), 200 mM NaCl and finally were purified by Mono Q column (GE Healthcare) using a gradient of 1 M NaCl. The purified AcrIIA4 was analyzed by SDS–polyacrylamide gel electrophoresis to confirm sample mass and sample purity (Figure 3)

3. 3. 2. *Spy*Cas9

The cells were harvested by centrifugation and resuspended in 20 mM Tris, pH 7.5, 300 mM NaCl, 5 mM BME, and 10% (w/v) glycerol. After disrupted by emulsiflex (Avestin, Canada) and centrifugation, the supernatant was loaded onto a 5 mL HisTrap HP column (GE Healthcare) pre-equilibrated with 20 mM Tris, pH 7.4, 300 mM NaCl, 5 mM BME, 10% (w/v) glycerol, and 30 mM imidazole.

*Spy*Cas9 was eluted with a linear gradient of 500 mM imidazole and then dialyzed against 20 mM Tris-HCl, pH 7.4, 175 mM NaCl, 10% (w/v) glycerol, and 5 mM BME. Dialyzed *Spy*Cas9 was loaded onto a 5 mL HiTrap SP column (GE Healthcare) pre-equilibrated with the dialysis buffer. *Spy*Cas9 was eluted with a linear gradient of 1 M NaCl. The protein was further purified using a HiLoad 16/600 Superdex200 column (GE Healthcare) equilibrated with 20 mM Tris-HCl pH 8.0, 200 mM NaCl, 5 % (w/v) glycerol, and 2 mM dithiothreitol (DTT). The purified AcrIIA4 was analyzed by SDS–polyacrylamide gel electrophoresis to confirm sample mass and sample purity (Figure 4)

3. 3. 3. *in vitro* transcription for preparation of sgRNA

The DNA template for RNA transcription was prepared by using the Giga-prep kit (Epigentics). The sgRNA was prepared *in vitro* by mixing rNTPs (rATP, rGTP, rCTP, and cUTP), $MgCl_2$, T7 RNA polymerase (P266L mutant)[57], inorganic pyrophosphatase, and the DNA template in the transcription buffer 50 mM Tris-HCl, pH 7.5, 15 mM $MgCl_2$, 5 mM DTT, 2 mM spermidine. After 6 hour transcription at 37°C, the synthesized RNA was precipitated by ethanol treatment overnight, purified using 12% denaturing polyacryl amide gel electrophoresis (19:1 cross-linking ratio), and electro-eluted (Elutrap, Whatman). The purified RNA was desalted and exchanged into water using Amicon (Millipore). The sequence of sgRNA is listed in Table 2.

Figure 3. Purification procedure of AcrIIA4.

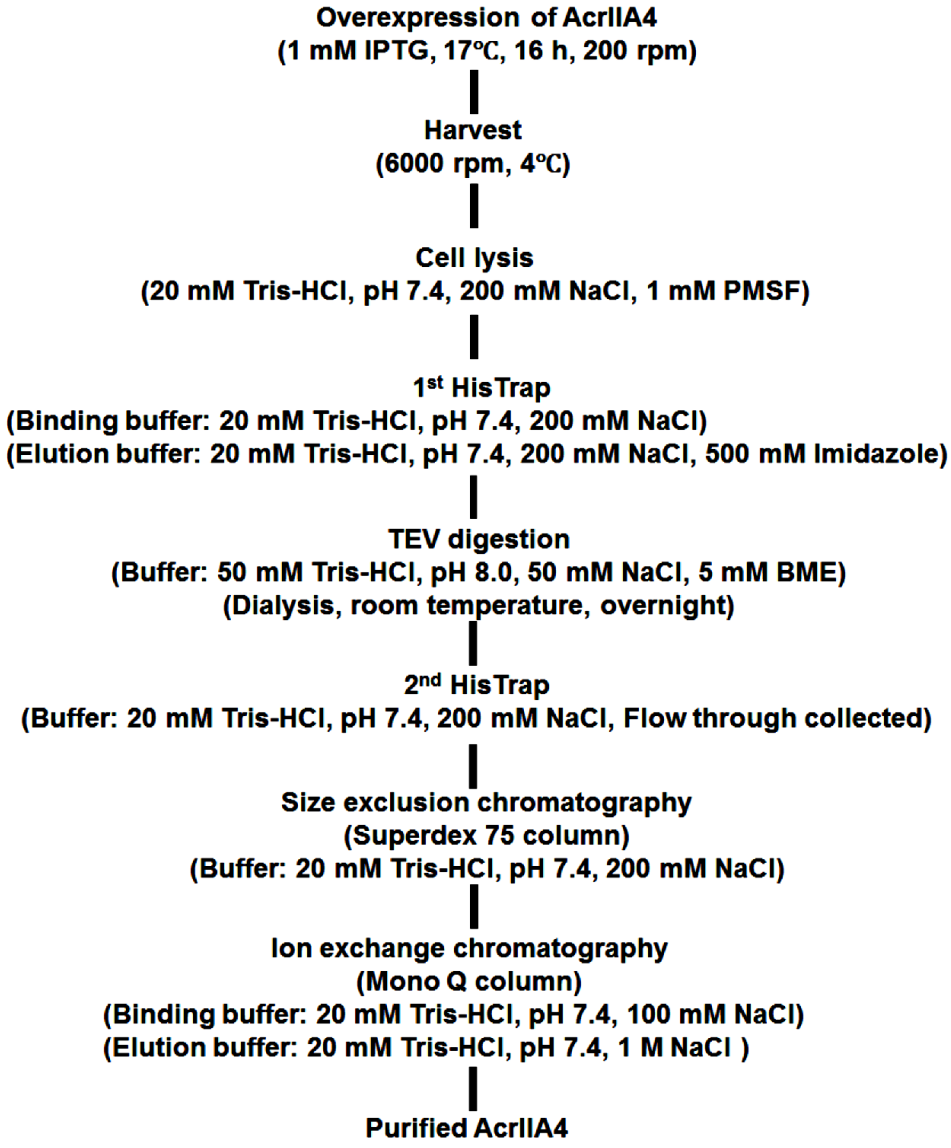

Overexpression of AcrIIA4
(1 mM IPTG, 17℃, 16 h, 200 rpm)
Harvest
(6000 rpm, 4℃)
Cell lysis
(20 mM Tris-HCl, pH 7.4, 200 mM NaCl, 1 mM PMSF)
1st HisTrap
(Binding buffer: 20 mM Tris-HCl, pH 7.4, 200 mM NaCl)
(Elution buffer: 20 mM Tris-HCl, pH 7.4, 200 mM NaCl, 500 mM Imidazole)
TEV digestion
(Buffer: 50 mM Tris-HCl, pH 8.0, 50 mM NaCl, 5 mM BME)
(Dialysis, room temperature, overnight)
2nd HisTrap
(Buffer: 20 mM Tris-HCl, pH 7.4, 200 mM NaCl, Flow through collected)
Size exclusion chromatography
(Superdex 75 column)
(Buffer: 20 mM Tris-HCl, pH 7.4, 200 mM NaCl)
Ion exchange chromatography
(Mono Q column)
(Binding buffer: 20 mM Tris-HCl, pH 7.4, 100 mM NaCl)
(Elution buffer: 20 mM Tris-HCl, pH 7.4, 1 M NaCl)
Purified AcrIIA4

Figure 4. Purification procedure of *Spy*Cas9

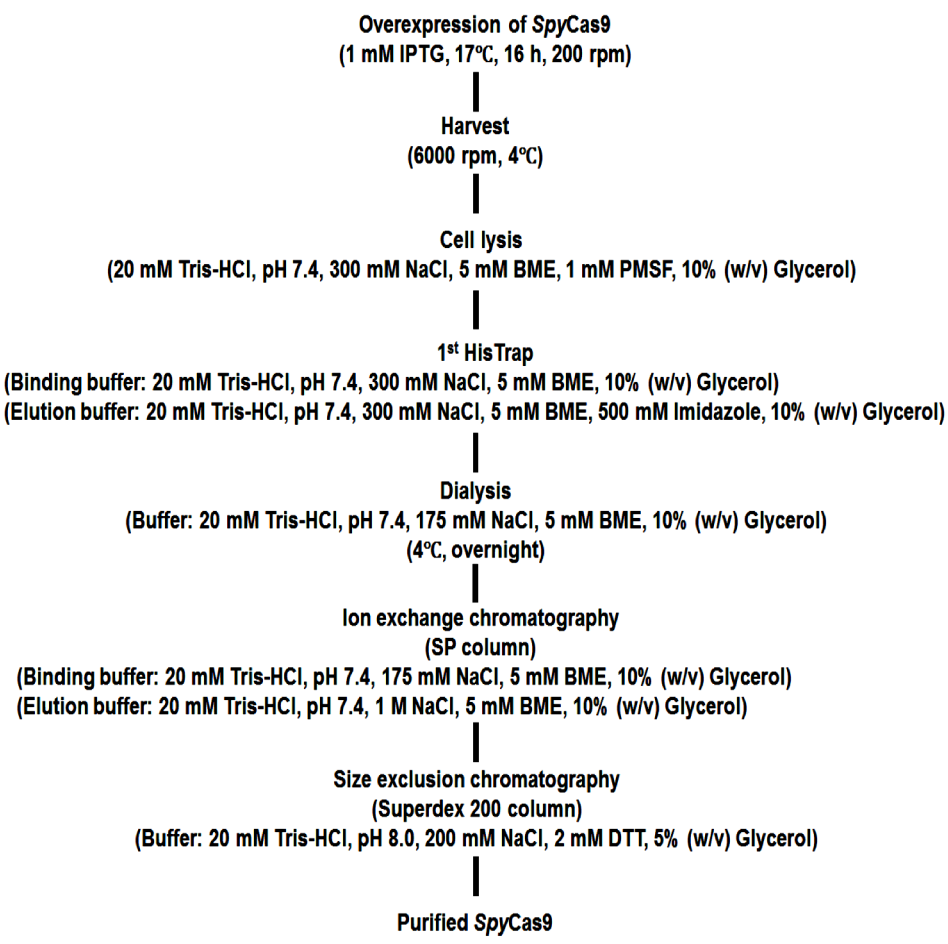

Overexpression of *Spy*Cas9
(1 mM IPTG, 17℃, 16 h, 200 rpm)
Harvest
(6000 rpm, 4℃)
Cell lysis
(20 mM Tris-HCl, pH 7.4, 300 mM NaCl, 5 mM BME, 1 mM PMSF, 10% (w/v) Glycerol)
1st HisTrap
(Binding buffer: 20 mM Tris-HCl, pH 7.4, 300 mM NaCl, 5 mM BME, 10% (w/v) Glycerol)
(Elution buffer: 20 mM Tris-HCl, pH 7.4, 300 mM NaCl, 5 mM BME, 500 mM Imidazole, 10% (w/v) Glycerol)
Dialysis
(Buffer: 20 mM Tris-HCl, pH 7.4, 175 mM NaCl, 5 mM BME, 10% (w/v) Glycerol)
(4℃, overnight)
Ion exchange chromatography
(SP column)
(Binding buffer: 20 mM Tris-HCl, pH 7.4, 175 mM NaCl, 5 mM BME, 10% (w/v) Glycerol)
(Elution buffer: 20 mM Tris-HCl, pH 7.4, 1 M NaCl, 5 mM BME, 10% (w/v) Glycerol)
Size exclusion chromatography
(Superdex 200 column)
(Buffer: 20 mM Tris-HCl, pH 8.0, 200 mM NaCl, 2 mM DTT, 5% (w/v) Glycerol)
Purified *Spy*Cas9

Table 2. Sequence of sgRNA

Description	Sequence (5’ to 3’)
sgRNA	GGA AAU UAG GUG CGC UUG GCG UUU UAG AGC UAG AAA UAG CAA GUU AAA AUA AGG CUA GUC CGU UAU CAA CUU GAA AAA GUG GCA CCG AGU CGG UGC UU

3. 4. NMR spectroscopy.

The NMR sample contained 2 mM ^{13}C,^{15}N-AcrIIA4 in 20 mM sodium phosphate, (pH 7.4), 100 mM NaCl, and 0.01% NaN_3. NMR spectra were recorded at 25°C on Bruker 600 MHz and 800 MHz spectrometers equipped with a *z*-shielded gradient triple resonance probe. Sequential and side chain assignments of ^{1}H, ^{15}N and ^{13}C resonances were achieved by three-dimensional triple resonance through-bond scalar correlation experiments CBCA(CO)NH, HNCACB, HNCO, HN(CA)CO, HBHA(CO)NH, HCCH-TOCSY, and ^{15}N-TOCSY-HSQC. 3D ^{13}C-separated NOESY and ^{15}N-separated NOESY experiments were obtained using a mixing time of 120 ms. The chi1 angle restraints were obtained based on the quantification of intraresidue NOEs between backbone amide proton, α proton, and β methylene protons[58]. Residual $^{1}D_{NH}$ dipolar couplings were obtained by taking the difference in the J splitting values measured in oriented (10 mg/ml of *pf1* phage alignment media) and isotropic (water) AcrIIA4 using 2D in–phase/antiphase ^{1}H– ^{15}N HSQC spectra[59]. ^{15}N-R_1 and ^{15}N-$R_1\rho$ relaxation rates, and heteronuclear ^{1}H– ^{15}N NOE were measured using pulse schemes described previously[60,61] on the Bruker 800 MHz spectrometer. Delays of 10, 50, 100, 400, 800, 1200, 1500 ms were used for the R_1 relaxation measurement, and 0, 4, 12, 24, 40, 60, 80, 96 ms used for $R_1\rho$ experiment. ^{15}N-R_2 relaxation rates were obtained from the ^{15}N-R_1 and ^{15}N-$R_1\rho$ relaxation rates using the equation, $R_2 = (R_1\rho - R_1 \sin^2 \theta)/\cos^2 \theta$, where $\theta = \arctan(\Omega_N/\gamma_N B_1)$, Ω_N is the resonance offset, and $\gamma_N B_1$ is the spin-lock field strength[62]. For the heteronuclear ^{1}H–^{15}N NOE measurement, ^{1}H saturation was achieved using 3 s of 120º ^{1}H pulses separated by 5 ms intervals. 2 s of relaxation delays were used for ^{15}N-R_1 and ^{15}N-$R_1\rho$ relaxation measurements, and 4s of the relaxation delay was used for the

heteronuclear $^{1}H-^{15}N$ NOE measurement. NMR spectra were processed using the NMRPipe[35], and analyzed using PIPP[36], and NMRView[37] programs.

3. 5. Structure calculations.

Interproton distance restraints were derived from the NOE spectra and classified into distance ranges according to the peak intensity. ϕ/ψ torsion angle restraints were derived from backbone chemical shifts using the program TALOS+[63]. The solution structure of AcrIIA4 was calculated by simulated annealing in torsion angle space using the Xplor-NIH program[64]. The target function of simulated annealing included a covalent geometry, a quadratic van der Waals repulsion potential[65], square-well potentials for interproton distance and torsion angle restraints, hydrogen bonding, RDC restraints[66], harmonic potentials for $^{13}C\alpha/^{13}C\beta$ chemical shift restraints[67], a multidimensional torsion angle database potential of mean force[68], and a radius of gyration term[69]. The radius of gyration represents a weak overall packing potential, and residues 1− 86 of AcrIIA4 were selected for the term. Structures were displayed using VMD-XPLOR software[70].

3. 6. Isothermal titration calorimetry (ITC)

The ITC experiment was performed at 25°C using an iTC200 calorimeter (Malvern). To characterize the affinity of AcrIIA4 to *Spy*Cas9, 50 μM *Spy*Cas9 in the cell was titrated with 500 μM of the AcrIIA4 in 20 mM sodium phosphate, (pH 7.4), 100 mM NaCl, 5 mM BME, and 0.01% NaN_3. For the measurement of affinity of AcrIIA4 to sgRNA-bound *Spy*Cas9, 10 μM *Spy*Cas9 with 12 μM sgRNA in the cell was titrated with 100 μM of AcrIIA4 in the same buffer. 50 μM sgRNA in the

cell was titrated with 500 μM of AcrIIA4 in the same buffer. Twenty consecutive 2 μL aliquots of AcrIIA4 were titrated into the cell and the duration of each injection was 4 sec, with injections occurring at intervals of 150 sec. The heats associated with dilution of the substrates were subtracted from the measured heats of binding. ITC titration data were analyzed using the Origin version 7.0 program provided with the instrument.

4. Results

4. 1. Resonance assignment of AcrIIA4

AcrIIA4 consists of 87 amino acid with molecular weight of 10.2 kDa. AcrIIA4 exists as a monomer in solution, and exhibits a well resolved ^{1}H-^{15}N HSQC NMR spectrum that is typically observed in compact folded proteins (Fig. 5 and 6).

NMR backbone assignment of AcrIIA4

Backbone assignment of AcrIIA4 for amide resonances obtained for 81 residues out of 87 residues (93%) was carried out using a suite of heteronuclear triple resonance NMR spectroscopy. I completely assigned backbone chemical shifts except for Thr28 that did not show backbone amide resonance in the 2D ^{1}H–^{15}N HSQC spectrum. Ser20, Ser24, Gln29, Glu47, and Gln65 did not show backbone amide resonances, but their aliphatic and carbonyl groups could be assigned from their sequential connectivities. ^{1}H, ^{15}N and ^{13}C assignment obtained for AcrIIA4 are listed in Table 3.

Figure 5. Size exclusion chromatogram of AcrIIA4.

Elution profiles of standard marker proteins are shown at the top of the chromatogram as a reference on Superdex 75 26/600 PG column (GE healthcare).

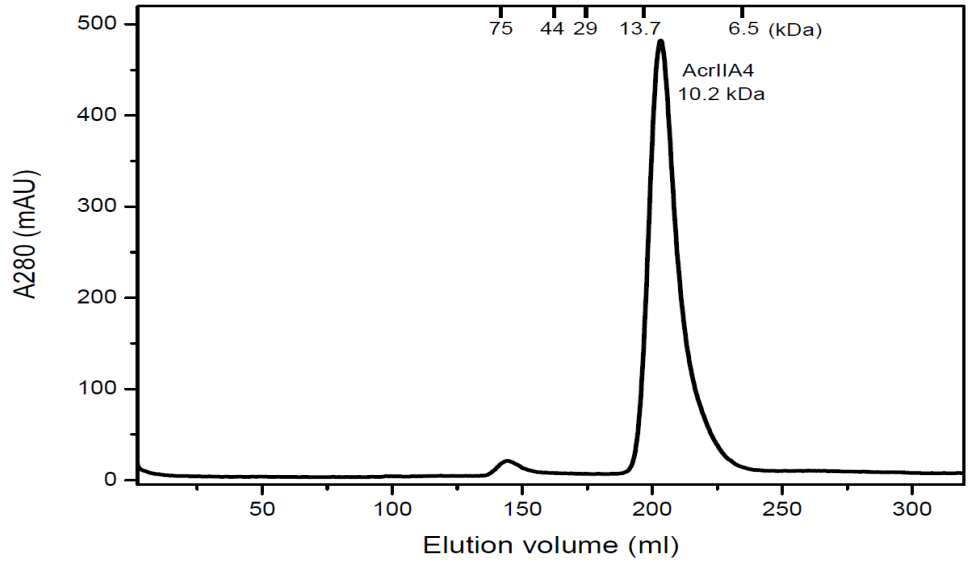

75
44
29
13.7
6.5 (kDa)
AcrIIA4
10.2 kDa
A280 (mAU)
500
400
300
200
100
0
50
100
150
200
250
300
Elution volume (ml)

Figure 6. ^{1}H-^{15}N HSQC spectra of AcrIIA4.

2 mM ^{13}C,^{15}N-labeled AcrIIA4 in 20 mM Sodium phosphate, pH 7.4, 100 mM NaCl, and 0.01% NaN_3 at 25°C. The backbone amide are annotated with the residue types and numbers.

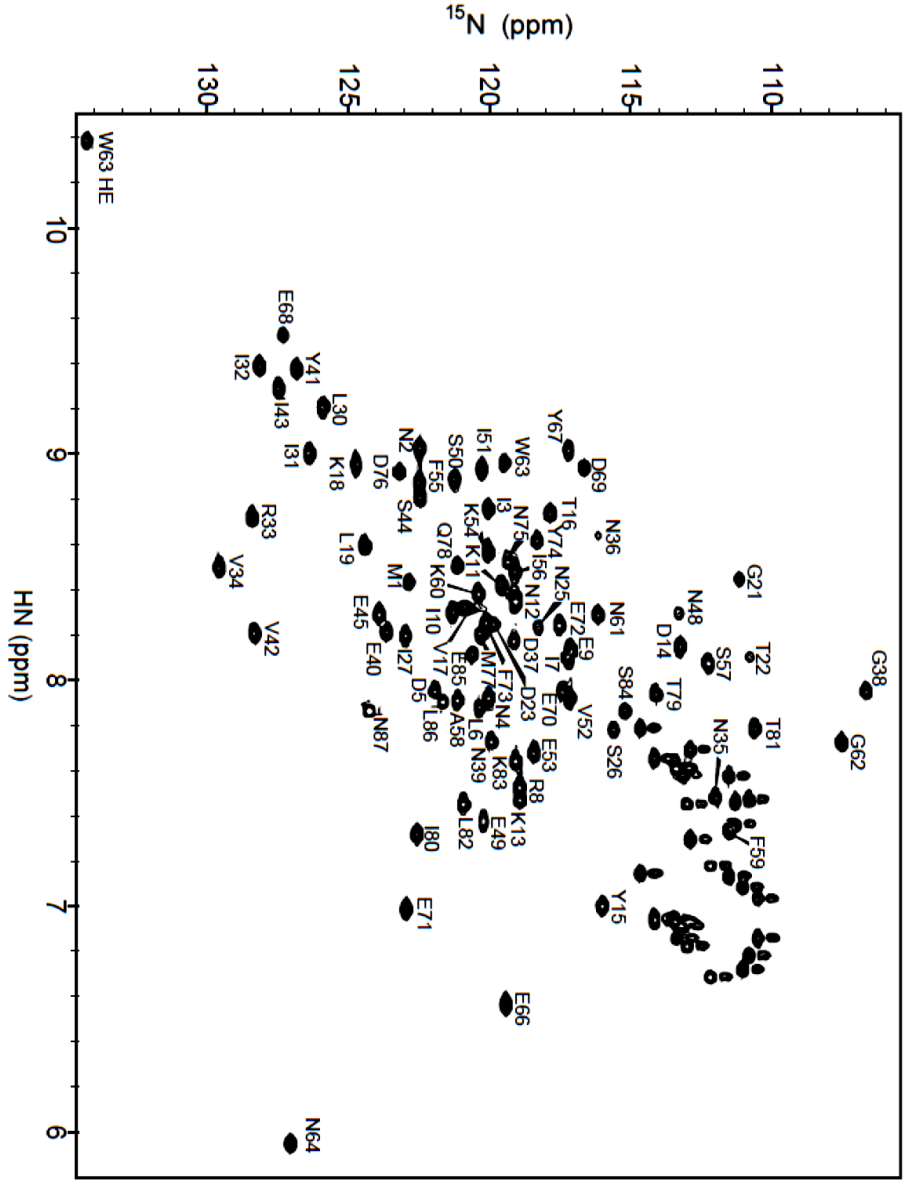

15N (ppm)
130
125
120
115
110
HN (ppm)
10
9
8
7
6
W63 HE
E68
Y41
I32
I43
L30
I31
K18
R33
V34
V42
D76
N2
S44
L19
M1
E45
E40
I27
N87
S50
F55
Q78
K60
I10
V17
D5
E85
L86
A58
I51
I3
K54
K11
W63
N75
I56
N12
D37
D23
M77
F73
N4
L6
N39
K83
R8
K13
E49
L82
I80
E71
E66
Y67
T16
Y74
N25
E72
I7
E70
V52
D69
N36
E9
N61
S84
S26
N48
D14
T79
G21
S57
N35
T22
T81
G62
G38
F59
Y15
N64

Table 3. Backbone ^{1}H, ^{15}N, $^{13}C\alpha$, $^{1}H_{\alpha}$, ^{13}CO and $^{13}C_{\beta}$, $^{1}H_{\beta}$ chemical shifts of AcrIIA4. (unit: ppm)

Residue	^{1}H	^{15}N	$^{13}C_{\alpha}$	$^{13}C_{\beta}$	^{13}CO	$^{1}H_{\alpha}$	$^{1}H_{\beta}$
M1	8.43	122.84	57.01	35.64	175.04	4.46	2.00
N2	9.03	122.46	52.28	41.00	176.69	4.92	3.15
I3	8.76	120.04	63.56	38.56	175.36	3.80	1.94
N4	7.92	120.02	57.28	38.57	178.56	4.35	2.78
D5	7.95	121.93	57.82	41.97	177.24	4.30	2.88
L6	7.88	120.34	58.99	42.04	177.86	3.74	1.82
I7	8.10	117.22	67.08	38.88	177.08	3.30	1.78
R8	7.52	118.93	60.46	30.48	178.90	3.86	2.04
E9	8.13	117.09	59.45	30.14	179.73	4.01	1.92
I10	8.30	121.32	67.39	39.73	179.08	3.37	1.68
K11	8.42	119.58	60.58	32.71	181.61	4.43	1.87
N12	8.34	119.06	55.59	38.83	176.64	4.60	2.92
K13	7.47	118.92	55.58	32.00	174.12	4.43	2.27
D14	8.15	113.25	56.74	39.18	174.85	4.23	2.85
Y15	7.00	116.00	58.84	40.49	176.31	4.66	2.78
T16	8.74	117.85	63.62	70.64	173.11	4.39	4.09
V17	8.32	120.87	59.46	35.71	174.05	5.33	1.90
K18	8.95	124.73	55.76	36.55	173.90	4.61	1.76
L19	8.59	124.41	54.22	45.08	176.77	5.25	1.62
S20						4.77	3.75
G21	8.45	111.17	46.66		174.37	4.27	
T22	8.10	110.80	63.32	70.21	175.09	4.25	4.33
D23	8.24	119.81	54.95	41.99	176.70	4.55	2.76
S24						4.33	3.90
N25	8.23	118.26	54.33	39.71	174.77	4.75	2.89
S26	7.78	115.60	58.62	65.00	173.34	4.56	3.81
I27	8.20	122.96	61.69	39.42	176.01	4.22	1.78
T28							

Residue	^{1}H	^{15}N	$^{13}C_{\alpha}$	$^{13}C_{\beta}$	^{13}CO	$^{1}H_{\alpha}$	$^{1}H_{\beta}$
Q29						5.17	1.85
L30	9.21	125.88	54.51	46.96	174.05	4.93	1.89
I31	9.00	126.36	61.68	40.70	175.48	4.79	1.72
I32	9.39	128.13	61.43	41.88	175.48	4.52	1.88
R33	8.72	128.39	57.44	30.32	175.39	4.44	1.96
V34	8.50	129.56	65.53	33.05	175.77	3.32	1.41
N35	7.48	112.00	52.49	40.58	175.21	4.72	3.22
N36	8.64	116.15	54.83	39.13	175.46	4.76	2.92
D37	8.17	119.14	55.23	42.12	176.54	4.71	2.75
G38	7.95	106.68	46.05		174.12	4.20	
N39	7.73	119.93	54.46	39.02	174.47	4.59	2.62
E40	8.21	123.62	55.45	33.90	174.18	4.99	1.72
Y41	9.37	126.82	58.65	42.15	174.29	4.49	3.05
V42	8.21	128.30	62.53	34.32	175.50	4.76	1.79
I43	9.29	127.45	61.65	40.90	174.61	4.35	1.91
S44	8.81	122.44	57.64	65.61	173.78	5.17	3.81
E45	8.29	123.90	58.12	31.51	175.62	4.11	2.12
S46	8.36	119.16	58.11	66.93	173.93	4.90	4.11
E47						4.05	2.01
N48	8.29	113.28	54.07	40.38	174.50	4.77	2.80
E49	7.37	120.21	57.04	34.09	173.87	4.58	1.85
S50	8.89	121.23	57.79	66.75	174.73	4.66	4.04
I51	8.94	120.27	66.43	38.43	177.58	3.50	1.74
V52	7.92	117.14	68.01	32.63	177.47	3.43	1.89
E53	7.67	118.42	60.65	30.32	180.17	3.89	2.27
K54	8.56	120.04	60.58	34.05	179.29	3.95	1.87
F55	8.87	122.47	62.98	40.36	176.38	3.87	3.56
I56	8.48	119.09	66.00	38.52	177.97	3.61	1.98
S57	8.07	112.25	62.37	63.77	177.97	4.00	3.91
A58	7.91	121.13	54.70	18.61	180.48	3.88	0.57

Residue	^{1}H	^{15}N	$^{13}C_{\alpha}$	$^{13}C_{\beta}$	^{13}CO	$^{1}H_{\alpha}$	$^{1}H_{\beta}$
F59	7.33	111.51	59.42	40.77	178.29	4.76	2.49
K60	8.38	120.39	60.41	31.66	177.23	4.42	1.98
N61	8.29	116.16	53.90	39.74	175.84	4.90	2.93
G62	7.72	107.54	45.94		173.62	4.26	
W63	8.96	119.46	59.27	29.12	175.41	4.11	3.44
N64	5.95	127.04	53.22	38.17	174.41	4.20	3.05
Q65			53.22	38.17	174.41	0.40	1.63
E66	6.56	119.42	60.74	30.69	176.55	3.64	1.62
Y67	9.02	117.21	57.53	39.55	177.64	5.06	3.32
E68	9.53	127.30	60.99	30.46	178.09	4.16	2.14
D69	8.94	116.65	55.95	40.32	176.81	4.56	2.76
E70	7.95	117.42	56.61	32.05	175.83	4.63	2.91
E71	6.98	122.94	61.13	29.61	176.85	2.34	-0.1
E72	8.24	117.52	60.82	29.71	177.94	3.88	2.02
F73	8.25	120.11	61.00	39.86	177.27	4.61	3.45
Y74	8.61	118.31	63.49	39.51	179.42	3.29	3.28
N75	8.52	119.35	56.93	38.09	178.59	4.55	3.01
D76	8.92	123.17	58.20	41.05	179.27	4.46	3.00
M77	8.20	120.33	57.52	34.06	179.77	4.06	0.17
Q78	8.50	121.13	60.03	28.58	180.27	3.86	2.48
T79	7.94	114.11	67.31	69.41	176.61	3.90	4.36
I80	7.32	122.56	66.61	38.34	178.08	3.62	2.27
T81	7.78	110.61	66.58	69.54	176.05	3.98	4.30
L82	7.45	120.92	57.23	42.52	178.68	4.25	1.79
K83	7.64	119.07	58.24	32.74	177.61	4.20	1.89
S84	7.86	115.21	60.56	64.72	177.62	4.33	3.90
E85	8.11	120.61	57.51	30.72	176.37	4.31	2.10
L86	7.90	121.64	55.81	43.06	176.27	4.34	1.62
N87	7.86	124.25	54.94	41.45	179.43	4.43	2.65

4. 2. Determination of solution structure of AcrIIA4

I have determined the solution structure of AcrIIA4 using heteronuclear triple-resonance NMR spectroscopy, which were CBCACONH, HNCBAB, HBHA(CO)NH, HCCH-TOCSY, and ^{15}N-TOCSY-HSQC, using 2.0 mM AcrIIA4 samples to assign backbone and side chain. The ^{1}H assignment data covered 602 out of 646 ^{1}H atoms excluding those in exchangeable hydroxyl and amino groups, which amounts to the 93.2% completeness of ^{1}H assignment. Three-dimensional ^{13}C-separated NOE and ^{15}N-separated NOE restraints were employed for the structure calculation using the Xplor-NIH program[64]. Residual dipolar couplings (RDCs) were measured in 10 mg/ml of *pf1* phage alignment medium. The structure was determined using 1,782 NMR restraints including 1,504 experimental NOE-based distance restraints, 161 dihedral angle restraints, 79 backbone $^{1}D_{NH}$ RDC restraints, and 38 hydrogen bonding restraints in Table 4. The hydrogen bond restraints were introduced at the final stage of the structure calculation, based on the secondary structural information from NOE and chemical shift data. The hydrogen bond restraints were in perfect agreement with the short-range and medium range NOEs, and dihedral angle restraints. The AcrIIA4 is comprised of three antiparallel β strands (β1, 16–19; β2, 29–33; β3, 40–44) and three flanking α helices (α1, 2–13; α2, 50–59; α3, 68–85) in an α1–β1–β2–β3–α2–α3 order (Fig. 7A). In addition, a short 3_{10} helix (22–25) is formed in the loop connecting β1 and β2 strands. Superposition of the backbone atoms for the ensemble of the 20 lowest energy structures of AcrIIA4 demonstrates that overall secondary structures are well defined such that three helices are packed together upon a three-stranded β sheet (Fig. 7B).

The connecting loop regions are less well-defined, and the β1–β2 loop among others shows the largest backbone root-mean-square deviations (RMSD ~1.3Å).

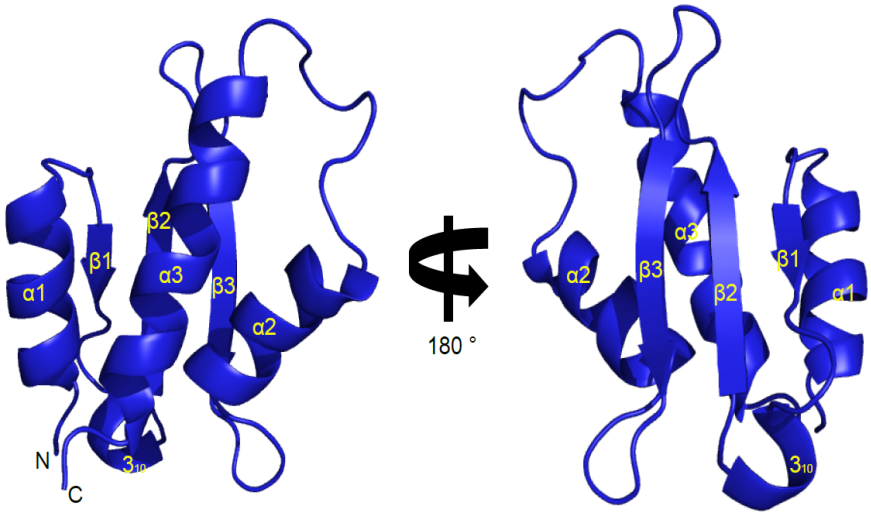
The overall backbone fold of free AcrIIA4 in solution was similar to that of the crystal structures determined for the AcrIIA4–*Spy*Cas9–sgRNA complexes, indicating that AcrIIA4 does not change the backbone conformation upon binding to *Spy*Cas9–sgRNA (Fig. 7C). The β1 strand was well-defined and the C-terminal α3 helix extended between Glu68 and Glu85 in the solution structure, which was consistent with the crystal structure. On the other hand, the β1–β2 loop formed a 3_{10} helix between Thr22 and Asn25 in the solution structure, but did not exhibit a regular secondary structure in complex with *Spy*Cas9–sgRNA.

AcrIIA4 features a cluster of aromatic residues that form a compact hydrophobic core via stacking and edge-to-face packing between the β3 strand, and α2 and α3 helices. The tightly packed aromatic side chains result in large ring current effects and unusual chemical shifts, such as Hβ of Glu71 (–0.1 ppm), Hα of Gln65 (0.4 ppm), and Hβ and Hγ of Met77 (0.17 ppm and –0.05 ppm).

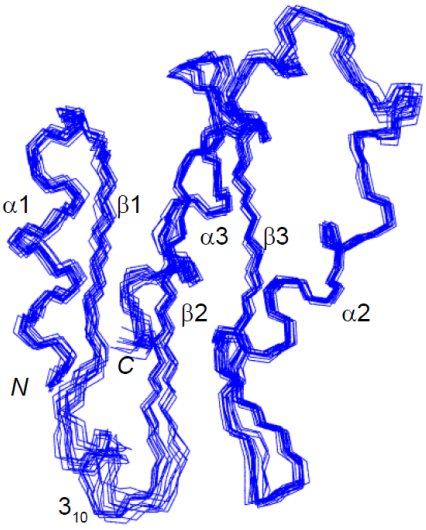
Figure 7. The structure of AcrIIA4 as ribbon representations.

Structures of AcrIIA4 as ribbon diagram. (A) Front and backside of AcrIIA4 in the free-state. Secondary structures, as well as N- and C-terminal are annotated. (B) Superposition of the backbone atoms of the final 20 simulated annealing structures of AcrIIA4. The structures are best-fit superposed against well-ordered secondary structures between residues 2-13, 16-19, 29-33, 40-44, 50-59, and 68-85. (C) AcrIIA4 from the crystal structure of the AcrIIA4-*Spy*Cas9-sgRNA complex (PDB code 5VW1).

A


β2
β1
α3
β3
α1
α2
N
C
3₁₀
180 °
α3
α2
β3
β1
β2
α1
3₁₀

B


α1
β1
α3
β3
β2
α2
C
N
3₁₀

C

Table 4. Restraints and structural statistics for AcrIIA4.

Experimental restraints	<SA>*
Nonredundant NOEs	1504
Dihedral angles, ϕ / ψ / χ	71 / 71 / 19
Hydrogen bonds	38
Residual dipolar coupling, $^1D_{NH}$	79
Total number of restraints	1782 (20.5 per residue)
rms deviation from experimental restraints	
Distances (Å) (1504)	0.029 ± 0.001
Torsion angles (°) (161)	0.90 ± 0.10
Residual dipolar coupling *R*-factor (%)†	
$^1D_{NH}$ (%) (79)	1.8 ± 0.2
rms deviation from idealized covalent geometry	
Bonds (Å)	0.002 ± 0
Angles (°)	0.47 ± 0.01
Impropers (°)	0.42 ± 0.01
Coordinate precision (Å)*‡	
Backbone	0.61 ± 0.11
Heavy atoms	1.38 ± 0.09
Ramachandran statistics (%)‡§	
Most favorable regions	91.2 ± 1.1
Allowed regions	8.8 ± 1.1

* For the ensemble of the final 20 simulated annealing structures

† The magnitudes of the axial and rhombic components of the alignment tensor were 7.3 Hz and 0.35, respectively.

‡ Regions with secondary structures (residues 2–13, 16–19, 29–33, 40–44, 50–59, and 68–85)

§ Calculated using the program PROCHECK[71]

4. 3. Backbone dynamics of AcrIIA4

Inherent dynamics of AcrIIA4

^{15}N R_1 and R_2 relaxation rates and ^{1}H–^{15}N heteronuclear NOE were measured to characterize the backbone dynamics of AcrIIA4. The rotational correlation time (τ_c) was calculated as 6.2 ns from the ratio of ^{15}N R_2 and R_1 relaxation rates in the ordered region, which is typical for the size of AcrIIA4 (~10.2 kDa). Secondary structural regions exhibited large ^{1}H–^{15}N heteronuclear NOE (> 0.8), and uniform ^{15}N R_2 and R_1 relaxation parameters indicating that the secondary structures were well structured and rigid without internal motions (Fig. 8). On the other hand, loop regions connecting the secondary structures manifested variable dynamic behavior in a wide range of timescales (Fig. 8). Reduced ^{1}H–^{15}N heteronuclear NOE (< 0.8) and ^{15}N R_1 relaxation rates indicated that fast backbone dynamics in ps–ns timescales was prevalent in the β1−β2 loop (Gly21−Ile27) region. In addition, increased ^{15}N R_2 relaxation rates indicated the presence of slower conformational exchanges in μs–ms timescales in the β1−β2 loop (Thr22, Asn23, and Ser26), the β3−α2 loop (Asn48 and Ser50) and the α2−α3 loop (Gly62 and Glu68) regions. The conformational exchanges also resulted in complete line broadening of backbone amide resonances in the β1−β2 loop (Ser20, Ser24, Thr28, and Gln29), the β3−α2 loop (Glu47), and the α2−α3 loop (Gln65) regions.

Relationship between the dynamic loop region and interfaces for Cas9

It is remarkable that the dynamic loop regions constitute the binding interfaces

between AcrIIA4 and *Spy*Cas9 in the crystal structure[50-52]. AcrIIA4 mainly interacts with the TOPO domain, the RuvC domain, and the C-terminal domain (CTD) of *Spy*Cas9. Specifically, AcrIIA4 employs the α1−β1 and β2−β3 loops to interact with the TOPO domain, the β1 strand and the β1−β2 loop to interact with the RuvC domain, and the β2−β3, β3−α2, and α2−α3 loops to interact with the CTD (Fig. 9B). The interfacial loops exhibited significant backbone dynamics in a wide range of time scales. The β1−β2 loop that binds to the RuvC domain of *Spy*Cas9 was highly mobile with fast internal motions in ps–ns timescales as well as slower conformational exchanges in μs–ms timescales (Fig. 9A). On the other hand, the β3−α2 loop and the α2−α3 loop that bind to the CTD of *Spy*Cas9 exhibited extensive conformation exchanges in μs–ms timescales. The binding interface for the TOPO domain appeared less dynamic compared to the other interfaces.

It has been reported that the fast and slow motions characterized by NMR relaxation play an important role in the target search and recognition process of biomolecular interactions[72-75]. It is notable that the dynamic β1−β2 loop of AcrIIA4 directly interacts with the active site of the *Spy*Cas9 nuclease, preventing the substrate DNA binding. *Spy*Cas9 cleaves double-stranded DNA substrates using two nuclease domains, the HNH nuclease domain that processes the target DNA strand complementary to the guide RNA, and the RuvC nuclease domain that processes the non-target DNA strand. The β1−β2 loop of AcrIIA4 extends to the active site of RuvC, and participates in a hydrogen bonding network with catalytic residues, blocking the access of the non-target DNA strand. Meanwhile, the α2−α3 loop that recognizes the PAM interaction site for the target DNA strand manifests slower

conformational exchanges. The α2–α3 loop does not directly block the active site of the nuclease, but occludes the binding site of the PAM sequence that is located immediately after the cleavage site of the target DNA strand. The loop dynamics of AcrIIA4 are thus linked to the target recognition and binding, with different time scales between the interaction surfaces.

AcrIIA4 is an acidic protein (pI ~4.2), and the surface electrostatic potential indicates that negative charges are densely populated in the β3–α2 loop, the α2–α3 loop and the beginning of the α3 helix (Fig. 9C). It has been suggested that the negatively charged patches mimic the phosphate groups of nucleic acids that associate with the PAM interaction site of *Spy*Cas9[50-52], which was also found in AcrF2 proteins inhibiting the subtype I-F effector complex[54].

Figure 8. Plots of relaxation parameters of backbone amide groups of AcrIIA4.

(A) ^{15}N R_1 relaxation (B) ^{15}N R_2 relaxation, and (C) ^{1}H-^{15}N heteronuclear NOE data as a function of residue number. The secondary structures of AcrIIA4 are shown on top as a schematic representation. The relaxation parameters were obtained on an 800 MHz Bruker NMR spectroscopy

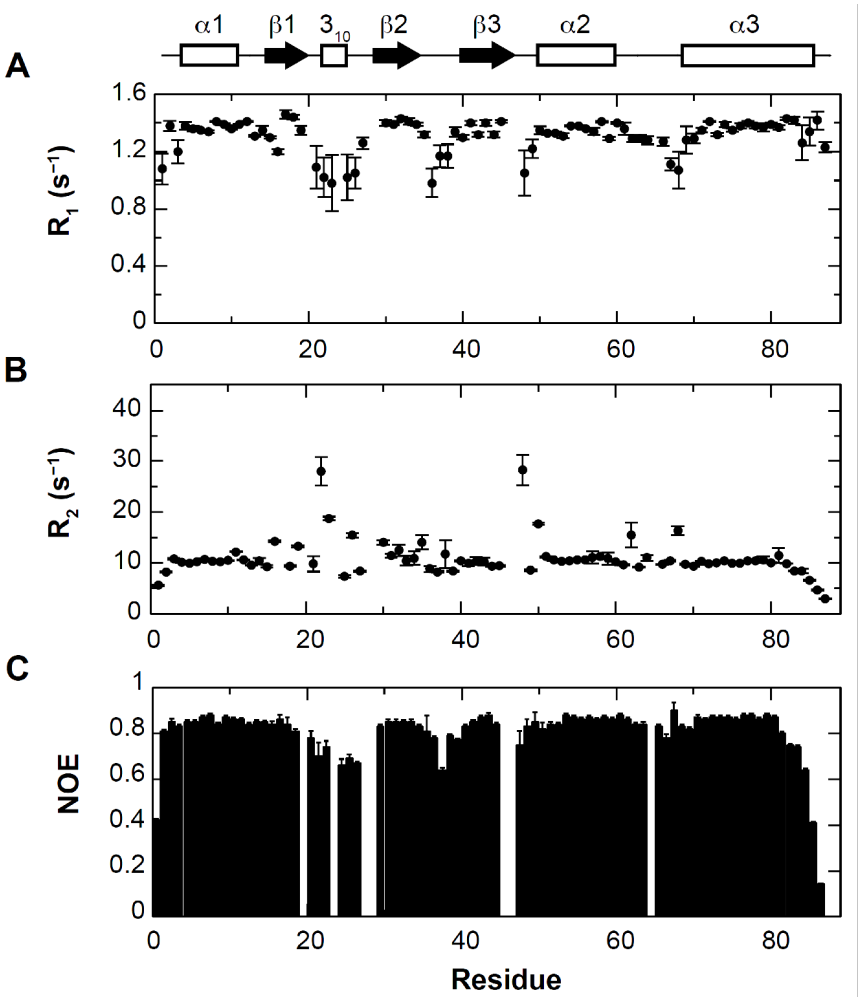
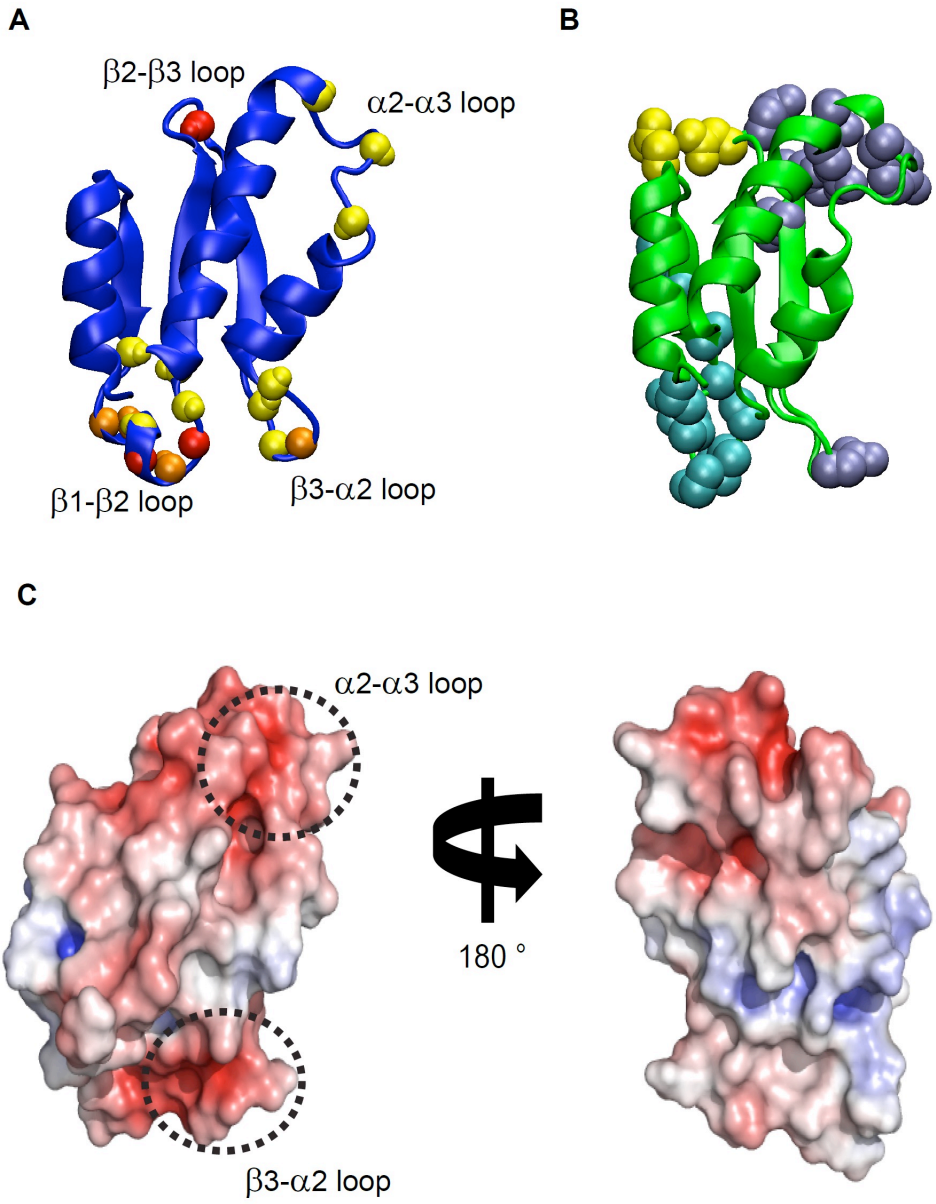

α1
β1
3₁₀
β2
β3
α2
α3
A
R1 (s−1)
1.6
1.2
0.8
0.4
0
0
20
40
60
80
B
R2 (s−1)
40
30
20
10
0
0
20
40
60
80
C
NOE
1
0.8
0.6
0.4
0.2
0
0
20
40
60
80
Residue

Figure 9. Structural relationship between dynamics of AcrIIA4 and interfaces for *Spy*Cas9

(A) Mobile regions with backbone dynamics of AcrIIA4 from the solution structure. AcrIIA4 is shown as a ribbon diagram, and backbone amide groups are shown as a space-filling model. Residues with fast internal motions (ps-ns time scale) and slower conformational dynamics (μs-ms) time scale) are colored in red and yellow, respectively. Residues with intermediate dynamics are colored in orange. (B) Binding interfaces of AcrIIA4 from the crystal structure in complex with *Spy*Cas9-sgRNA (PDB code 5VW1). AcrIIA4 is shown as a ribbon diagram, and interfacial residues in the complex are shown as a space filling model. Interfaces for the TOPO domain, the RuvC domain, and the C-terminal domain of *Spy*Cas9 are colored in yellow, cyan, and ice-blue, respectively. (C) Front and backside view of molecular surface representations of AcrIIA4 color-coded by electrostatic potential, ±5 kT. AcrIIA4 on the left panel is drawn in the same orientation as in (A).


A
β2-β3 loop
α2-α3 loop
β1-β2 loop
β3-α2 loop
B
C
α2-α3 loop
180 °
β3-α2 loop

4. 4. Interaction of AcrIIA4 to *Spy*Cas9

I quantitatively measured the binding thermodynamics between AcrIIA4 and apo-*Spy*Cas9, sgRNA-bound *Spy*Cas9 or sgRNA using isothermal titration calorimetry (ITC). In the interaction between AcrIIA4 and apo-*Spy*Cas9, the measurement revealed a large exothermic reaction between AcrIIA4 and *Spy*Cas9 with K_D = 4.8 ± 0.2 μM and 1:1 stoichiometry (Fig. 10A). The interaction between AcrIIA4 and sgRNA bound-*Spy*Cas9 showed a significantly higher affinity, with the K_D value of 0.6 ± 0.1 nM (Fig. 10B). The binding affinity measured by ITC was similar to the previous report from the microscale thermophoresis data[50]. Thus, the presence of sgRNA promoted the interaction between AcrIIA4 and *Spy*Cas9 by a factor of 8000 (Table 5). The weaker binding between AcrIIA4 and apo-*Spy*Cas9 explains previous observations that AcrIIA4 did not co-elute with *Spy*Cas9 in the absence of sgRNA from the size exclusion chromatography, though MBP pull-down assays showed a possible weak interaction between AcrIIA4 and apo-*Spy*Cas9[50-52]. I also note that AcrIIA4 did not exhibit measurable affinity to sgRNA alone (Fig. 10C). The binding of AcrIIA4 and apo-*Spy*Cas9 was largely an enthalpy-driven reaction which implicates their binding interfaces comprising extensive electrostatic interaction and hydrogen bonding networks[52]. The entropic contribution to the binding was unfavorable, which reflects a decrease in conformational flexibility at the interfaces upon complex formation. Indeed, the increased affinity of AcrIIA4 to sgRNA-bound *Spy*Cas9 largely originated from favorable entropic contribution, demonstrating that binding of sgRNA to *Spy*Cas9 relieved the entropic cost of *Spy*Cas9 to accommodate AcrIIA4.

^{1}H-^{15}N HSQC spectra of ^{15}N-AcrIIA4 titrated with unlabeled *Spy*Cas9 exhibited

gradual line broadening without chemical shift changes (Fig. 11). Due to the large molecular weight of *Spy*Cas9 (158.4 kDa), backbone amide resonances of free AcrIIA4 gradually disappeared without any sign of new signals from the complex, such that 0.2 mM ^{15}N-AcrIIA4 complexed with *Spy*Cas9 at a 1:1 molar ratio resulted in complete line broadening, except for the C-terminal Asn87. Most amide resonances of free AcrIIA4 exhibited intensity reduction of >60%, when the complex formation reached 20% from the stoichiometric titration. This indicates that the line broadening is not only caused by large molecular weight upon complex formation, but also by intermediate exchanges from the interaction. Those residues with larger line broadening during the titration include Asn36 that interacts with the TOPO domain of *Spy*Cas9, and Lys18, Gly21, and Thr22 that interact with RuvC nuclease domain. Asn48 and Tyr67 that interact with the CTD of *Spy*Cas9 also exhibited significant line broadening. Residues close to the binding interface such as Glu49, Ser50, Ile51, and Glu68 also showed large line broadening. Thus, the chemical shift perturbation data is fully consistent with the ITC results that AcrIIA4 associates with *Spy*Cas9 in the absence of sgRNA. Also, the residues with line broadening suggest that AcrIIA4 binds to apo-*Spy*Cas9 in a similar manner as the crystal structure of the ternary complex.

The crystal structure of the AcrIIA4-*Spy*Cas9-sgRNA complex has shown that *Spy*Cas9 employs the RuvC domain, TOPO domain, and CTD to interact with AcrIIA4[50-52]. The interfaces at the TOPO domain and the CTD (Fig. 12, red spheres) overlap with the PAM interaction site of *Spy*Cas9 for the target strand of substrate DNA. The interface at the RuvC domain (Fig. 12, orange sphere) is the active site pocket that cleaves the non-target strand of substrate DNA. When the binding

interfaces in the complex are mapped onto apo-*Spy*Cas9, interfaces from the RuvC domain and the CTD are relatively well maintained in apo-*Spy*Cas9, but those from the TOPO domain are completely lost due to high mobility Based on the structural alignment, I postulate that the RuvC domain and the CTD provide the interaction surface for AcrIIA4 in the absence of sgRNA. Binding of *Spy*Cas9 to sgRNA induces a major rearrangement of the helical REC domain, as well as ordering of the PAM interaction site at the TOPO domain and the CTD. It is remarkable that the affinity of the AcrIIA4-*Spy*Cas9 interaction increases by a thousand-fold upon sgRNA binding. Thus, the guide RNA binding and concomitant conformational rearrangement of *Spy*Cas9 provide key interfaces not only for the DNA substrate but also for the inhibitor AcrIIA4.

Figure 10. ITC for the interaction between AcrIIA4 and *Spy*Cas9, and between AcrIIA4 and sgRNA-bound *Spy*Cas9.

Interactions (A) between AcrIIA4 and apo-*Spy*Cas9, (B) between AcrIIA4 and sgRNA-bound *Spy*Cas9 and a titration between (C) AcrIIA4 and sgRNA. Raw ITC data (top panel) and integrated heats of injection (bottom panel) are presented for the titration between AcrIIA4 and *Spy*Cas9. In the bottom panel, the experimental data are shown as solid squares and the least squares best-fit curves derived from a simple one-site binding model are shown as a black line. The thermodynamic parameters for the complex formation are listed in Table 5. For the AcrIIA4 and sgRNA, the lack of characteristic binding isotherm indicates that AcrIIA4 did not bind to sgRNA alone.

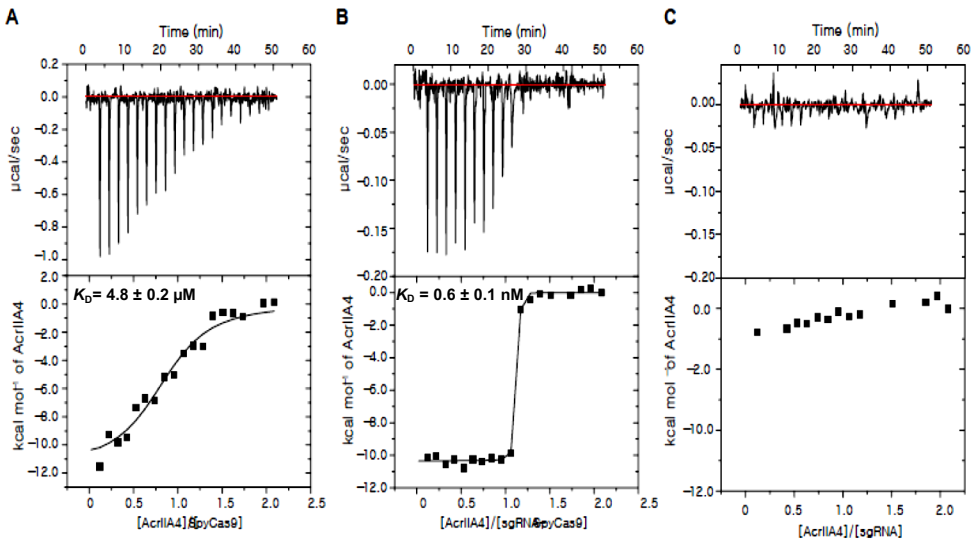

A
B
C
Time (min)
µcal/sec
kcal mol⁻¹ of AcrIIA4
K_D= 4.8 ± 0.2 µM
K_D = 0.6 ± 0.1 nM
[AcrIIA4]/[sgRNA:SpyCas9]
[AcrIIA4]/[sgRNA]

Figure 11. Superimposed ^{1}H-^{15}N HSQC spectra of AcrIIA4 titrated with unlabeled *Spy*Cas9.

0.2 mM AcrIIA4 in free-state is shown in black. The concentration of *Spy*Cas9 in the titration were 2 μM (red), 4 μM (orange), 10 μM (green), 40 uM (magenta). Residues at the binding interface for sgRNA-bound *Spy*Cas9 is annotated in red. Residues with large line broadening and located near the interface are annotated in black.

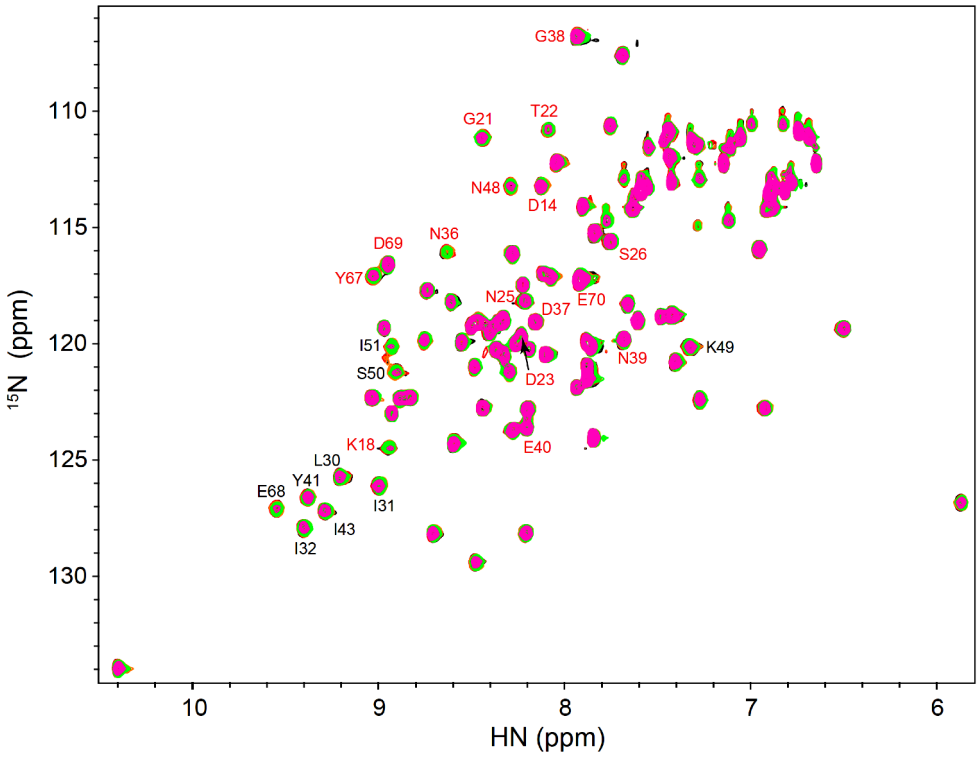

G38
G21
T22
N48
D14
D69
N36
S26
Y67
N25
D37
E70
I51
N39
K49
S50
D23
K18
E40
L30
Y41
E68
I31
I43
I32
110
115
120
125
130
^{15}N (ppm)
10
9
8
7
6
HN (ppm)

Figure 12. Structural comparison of binding interfaces of *Spy*Cas9 for AcrIIA4

Top panel: Domain organization of *Spy*Cas9 comprising the RuvC nuclease domain, the bridge helix (BH), the RNA recognition domain (REC), the HNH nuclease domain, the topoisomerase-homology domain (TOPO), and the C-terminal domain (CTD). Middle left panel: The crystal structure of apo-*Spy*Cas9 in a ribbon diagram. Individual domains of *Spy*Cas9 are colored according to the color scheme of domain organization. The interfacial residues of *Spy*Cas9 for AcrIIA4 are mapped onto the crystal structure of apo-*Spy*Cas9 (PDB code 4CMP) in a space filling model and PAM recognition site and RuvC nuclease site are colored in red and orange, respectively. Middle right panel: The crystal structure of the AcrIIA4-*Spy*Cas9-sgRNA complex (PDB code 5VW1) in a ribbon diagram. AcrIIA4 and sgRNA are colored in cyan and magenta, respectively. The interfacial residues and the domains are colored according to the same color scheme as in the apo-*Spy*Cas9 structure. The binding interfaces are enlarged in bottom left and right panel, respectively, and AcrIIA4 and sgRNA are omitted for visual clarity in the complex structure.

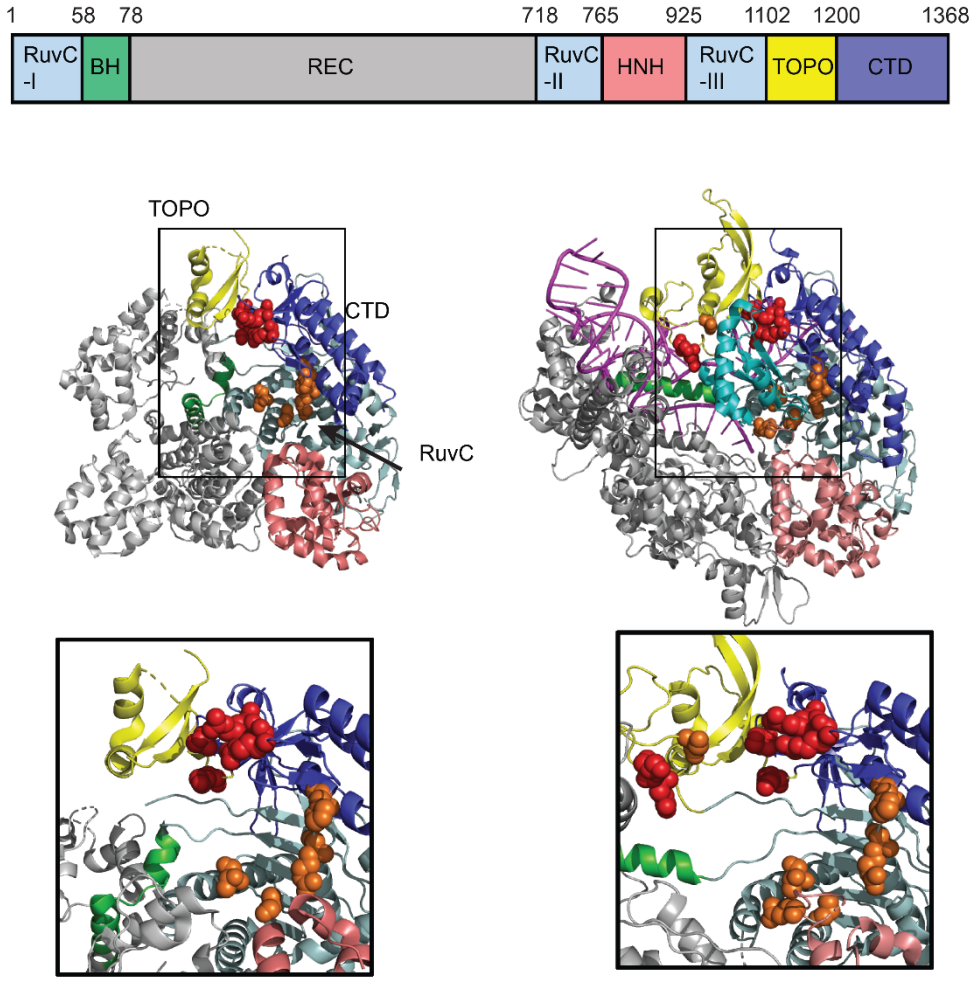

1
58
78
718
765
925
1102
1200
1368
RuvC -I
BH
REC
RuvC -II
HNH
RuvC -III
TOPO
CTD
TOPO
CTD
RuvC

Table 5. Thermodynamic parameters for the interaction between AcrIIA4 and *Spy*Cas9 obtained by isothermal titration calorimetry at 25°C.

Description	K_D (nM)	ΔG (kcal/mol)	ΔH (kcal/mol)	$-T\Delta S$ (kcal/mol)
AcrIIA4:*Spy*Cas9	4800 ± 150	− 7.3 ± 0.1	− 11.5 ± 0.7	4.2 ± 0.8
AcrIIA4:*Spy*Cas9–sgRNA	0.6 ± 0.1	− 12.6 ± 0.4	− 10.3 ± 0.1	− 2.3 ± 0.4

5. Discussion

In this study, I investigated the solution structure of AcrIIA4. The structure of AcrIIA4 shows distinct and compact monomeric αβββαα fold. The overall backbone fold of AcrIIA4 does not change the backbone formation upon binding to *Spy*Cas9-sgRNA except for 3_{10} helix within β1-β2 loop. Upon binding to *Spy*Cas9, the 3_{10} helix turns to protruding β1-β2 loop, thereby inserted into the binding interface of RuvC nuclease domain which is important for catalytic activity. The cluster of aromatic residues (Tyr41, Phe55, Phe59, Trp63, and Tyr67) forms a compact hydrophobic core by stacking and edge-to face packing, contributing to a relatively stable conformation. I speculate that the rigid body of AcrIIA4 and small conformational changes upon Cas9 binding may have evolved to provide a backbone scaffold and dynamic loops for proper binding and to mimic double stranded DNA conformation in the type II-A CRISPR-Cas9 system.

Furthermore, I quantitatively measured backbone dynamics of AcrIIA4, and it reveals characteristic motions at the binding interfaces for *Spy*Cas9. The fast and slow motions at the backbone of β1-β2 loop could contribute to interact with RuvC domain, and slower conformational exchange motions of β3-α2 loop and the α2-α3 loop could contribute to the interaction with CTD. The α1-β1 loop for TOPO domain relatively lacks dynamic motion. However, in the crystal structure of apo-SpyCas9, the residues corresponding to interfacial residues for AcrIIA4 in the AcrIIA4-SpyCas9-sgRNA complex lack the electron density map due to high mobility[76]. I

speculate that the binding interfaces with high mobility motion within TOPO domain in the apo-*Spy*Cas9 may contribute to recognition and interaction with AcrIIA4. It remains to figure out whether a mutation at the regions exhibiting dynamic backbone motion in the AcrIIA4 could increase the affinity and association rate to Cas9.

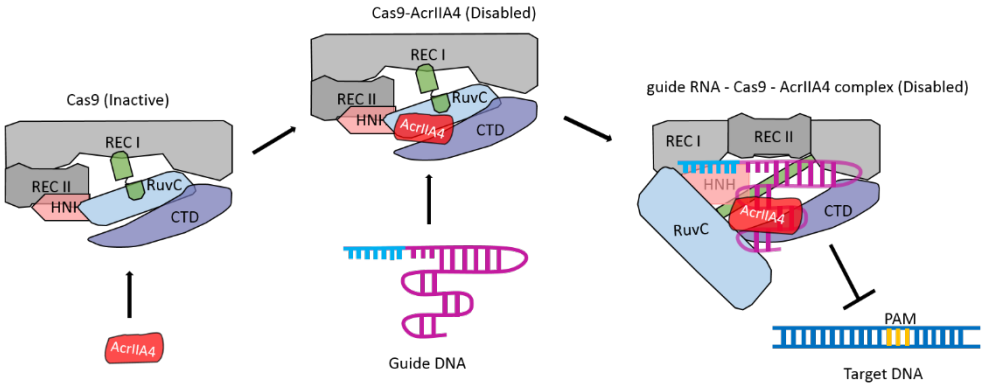
I quantitatively measured the binding thermodynamics between AcrIIA4 and *Spy*Cas9. In the presence of sgRNA, AcrIIA4 binds stronger to *Spy*Cas9. Since the interfaces for AcrIIA4 are relatively well maintained even in the apo-*Spy*Cas9, I speculate that REC lobe which suffers a large conformational change upon sgRNA-binding could be rather a steric hindrance to AcrIIA4 with moderate affinity. It has been shown that AcrIIA4 competes with the double-strand DNA substrate for *Spy*Cas9 binding with a ~ 17-fold higher affinity[50]. AcrIIA4 inhibited the substrate DNA binding to *Spy*Cas9 in a concentration-dependent manner, but did not affect the dissociation of the DNA substrate from *Spy*Cas9, indicating a slow dissociation kinetics[51]. In other words, AcrIIA4 effectively inhibits the nuclease activity of *Spy*Cas9, but there is competition between the DNA substrate and AcrIIA4 targeting for the same interface on *Spy*Cas9.

I speculate that AcrIIA4 bound sequentially to *Spy*Cas9 and then to guide RNA with increasing affinity (Fig.13). In the sequential binding mechanism, AcrIIA4 does not need to compete with target DNA for *Spy*Cas9 binding, since target DNA recognition requires a pre-formed assembly of guide RNA and *Spy*Cas9. Since an electrophoretic mobility shift assay showed AcrIIA4 does not affect sgRNA binding to *Spy*Cas9[51], AcrIIA4 is able to form a binary complex with apo-*Spy*Cas9 as well AcrIIA4-*Spy*Cas9-sgRNA complex. The binary complex has a moderate affinity, but it can rapidly turn into a tight ternary complex in the presence of matured guide RNA,

effectively preventing the target DNA binding. Since maturation of transactivating crRNA (tracrRNA) and pre-crRNA is mainly processed by RNase III, and Cas9 stabilize the interaction between tracrRNA and pre-crRNA for RNase III activity in streptococcus pyogenes and thermophilus type-II A system[78], it remains to be seen whether AcrIIA4 can not only inhibit Cas9 in target DNA binding but also affect function of Cas9 for the maturation of tracrRNA and crRNA. I speculate that *in vivo* AcrIIA4 can not affect the tracrRNA:pre-crRNA binding to Cas9 because the dual RNA accommodates in REC lobe and RuvC domain distal from the interface for AcrIIA4. Considering that a half-life of nucleic acids is longer than that of protein, the sequential binding mechanism would be advantageous for the anti-CRISPR function, especially when the anti-CRISPR genes are integrated as a prophage, and constitutively inactivates the host CRISPR-Cas system[77] This study provides insight into anti-CRISPR-mediated inhibition mechanisms for disabling the SpyCas9, thereby broadening CRISPR-Cas regulatory tools for genome editing.

Figure 13. Schematic illustration of a sequential binding mechanism for Cas9 inhibition

In the sequential binding mechanism, AcrIIA4 preoccupies the PAM recognition site in the CTD and the cleavage site on the RuvC domain with a moderate affinity, forming AcrIIA4-*Spy*Cas9 complex. When guide RNA is transcribed and close to AcrIIA4-SpyCas9 binary complex, the binary complex turns to a tight ternary complex, AcrIIA4-*Spy*Cas9-sgRNA complex, which inhibits target DNA recognition.


Cas9 (Inactive)
REC I
REC II
RuvC
HNK
CTD
AcrIIA4
Cas9-AcrIIA4 (Disabled)
REC I
REC II
RuvC
HNK
AcrIIA4
CTD
Guide DNA
guide RNA - Cas9 - AcrIIA4 complex (Disabled)
REC I
REC II
HNK
AcrIIA4
CTD
RuvC
PAM
Target DNA

6. Conclusion

In this paper, I have determined the solution structure of the anti-CRISPR protein AcrIIA4, and investigated its backbone dynamics using NMR spectroscopy. AcrIIA4 is a small monomeric protein and adopts a compact αβββαα fold, AcrIIA4 exhibits significant dynamics in the loop region that serve as the binding interface for *Spy*Cas9. Remarkably, AcrIIA4 can both bind to *Spy*Cas9 in the absence and presence of sgRNA, but interact with sgRNA-bound *Spy*Cas9 with stronger affinity by a thousand-fold. The binary complex between AcrIIA4 and *Spy*Cas9 can associate with sgRNA to form a tight ternary complex, avoiding competition with the DNA substrate for *Spy*Cas9 binding.

III. REFERENCES

1 Neeper, M. *et al.* Cloning and expression of a cell surface receptor for advanced glycosylation end products of proteins. *J. Biol. Chem.* **267**, 14998-15004 (1992).

2 Schmidt, A. M. *et al.* Isolation and characterization of two binding proteins for advanced glycosylation end products from bovine lung which are present on the endothelial cell surface. *J. Biol. Chem.* **267**, 14987-14997 (1992).

3 Hori, O. *et al.* The receptor for advanced glycation end products (RAGE) is a cellular binding site for amphoterin mediation of neurite outgrowth and co-expression of rage and amphoterin in the developing nervous system. *J. Biol. Chem.* **270**, 25752-25761 (1995).

4 Tian, J. *et al.* Toll-like receptor 9–dependent activation by DNA-containing immune complexes is mediated by HMGB1 and RAGE. *Nat. Immunol.* **8**, 487-496 (2007).

5 Orlova, V. V. *et al.* A novel pathway of HMGB1-mediated inflammatory cell recruitment that requires Mac-1-integrin. *The EMBO journal* **26**, 1129-1139 (2007).

6 Park, L. *et al.* Suppression of accelerated diabetic atherosclerosis by the soluble receptor for advanced glycation endproducts. *Nat. Med.* **4**, 1025-1031 (1998).

7 Yan, S. D. *et al.* RAGE: A potential target for Aβ-mediated cellular perturbation in Alzheimer's disease. *Curr. Mol. Med.* **7**, 735-742 (2007).

8 Taguchi, A. *et al.* Blockade of RAGE–amphoterin signalling suppresses tumour growth and metastases. *Nature* **405**, 354-360 (2000).

9 Fritz, G. RAGE: a single receptor fits multiple ligands. *Trends Biochem. Sci.* **36**, 625-632 (2011).

10 Yan, S. F., Ramasamy, R. & Schmidt, A. M. Receptor for AGE (RAGE) and its ligands—cast into leading roles in diabetes and the inflammatory response. *Journal of Molecular Medicine* **87**, 235-247 (2009).

11 Bierhaus, A. *et al.* Understanding RAGE, the receptor for advanced glycation end products. *Journal of Molecular Medicine* **83**, 876-886 (2005).

12 Sirois, C. M. *et al.* RAGE is a nucleic acid receptor that promotes inflammatory responses to DNA. *The Journal of experimental medicine* **210**, 2447-2463 (2013).

13 Schmidt, A. M., Du Yan, S., Yan, S. F. & Stern, D. M. The biology of the receptor for advanced glycation end products and its ligands. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research* **1498**, 99-111 (2000).

14 Donato, R. RAGE: a single receptor for several ligands and different cellular responses: the case of certain S100 proteins. *Curr. Mol. Med.* **7**, 711-724 (2007).

15 Heizmann, C., Ackermann, G. & Galichet, A. in *Calcium Signalling and Disease* 93-138 (Springer, 2007).

16 Donato, R. S100: a multigenic family of calcium-modulated proteins of the EF-hand type with intracellular and extracellular functional roles. *The international journal of biochemistry & cell biology* **33**, 637-668 (2001).

17 Huttunen, H. J., Fages, C. & Rauvala, H. Receptor for advanced glycation end products (RAGE)-mediated neurite outgrowth and activation of NF-κB require the cytoplasmic domain of the receptor but different downstream signaling pathways. *J. Biol. Chem.* **274**, 19919-19924 (1999).

18 Schäfer, B. W., Wicki, R., Engelkamp, D., Mattei, M.-g. & Heizmann, C. W. Isolation of a YAC clone covering a cluster of nine S100 genes on human chromosome 1q21: rationale for a new nomenclature of the S100 calcium-binding protein family. *Genomics* **25**, 638-643 (1995).

19 Marenholz, I., Heizmann, C. W. & Fritz, G. S100 proteins in mouse and man: from evolution to function and pathology (including an update of the nomenclature). *Biochem. Biophys. Res. Commun.* **322**, 1111-1122 (2004).

20 Schafer, B. W. *et al.* Brain S100A5 is a novel calcium-, zinc-, and copper ion-binding protein of the EF-hand superfamily. *J. Biol. Chem.* **275**, 30623-30630, doi:10.1074/jbc.M002260200 (2000).
21 Chan, W. Y., Xia, C. L., Dong, D. C., Heizmann, C. W. & Yew, D. T. Differential expression of S100 proteins in the developing human hippocampus and temporal cortex. *Microsc. Res. Tech.* **60**, 600-613 (2003).
22 Hancq, S. *et al.* S100A5: a marker of recurrence in WHO grade I meningiomas. *Neuropathol. Appl. Neurobiol.* **30**, 178-187 (2004).
23 Teratani, T. *et al.* Restricted expression of calcium-binding protein S100A5 in human kidney. *Biochem. Biophys. Res. Commun.* **291**, 623-627 (2002).
24 Bertini, I. *et al.* Solution structure and dynamics of S100A5 in the apo and Ca2+-bound states. *Journal of biological inorganic chemistry : JBIC : a publication of the Society of Biological Inorganic Chemistry* **14**, 1097-1107, doi:10.1007/s00775-009-0553-1 (2009).
25 Nishikawa, T. *et al.* Identification of S100b protein as copper-binding protein and its suppression of copper-induced cell damage. *J. Biol. Chem.* **272**, 23037-23041 (1997).
26 Wheeler, L. C. & Harms, M. J. Human S100A5 binds Ca 2+ and Cu 2+ independently. *BMC Biophysics* **10**, 8 (2017).
27 Leclerc, E., Fritz, G., Vetter, S. W. & Heizmann, C. W. Binding of S100 proteins to RAGE: an update. *Biochim. Biophys. Acta* **1793**, 993-1007, doi:10.1016/j.bbamcr.2008.11.016 (2009).
28 Xie, J. *et al.* Hexameric calgranulin C (S100A12) binds to the receptor for advanced glycated end products (RAGE) using symmetric hydrophobic target-binding patches. *J. Biol. Chem.* **282**, 4218-4231 (2007).
29 Ostendorp, T. *et al.* Structural and functional insights into RAGE activation by multimeric S100B. *The EMBO journal* **26**, 3868-3878 (2007).
30 Leclerc, E., Fritz, G., Weibel, M., Heizmann, C. W. & Galichet, A. S100B and S100A6 differentially modulate cell survival by interacting with distinct RAGE (receptor for advanced glycation end products) immunoglobulin domains. *J. Biol. Chem.* **282**, 31317-31331 (2007).
31 Koch, M. *et al.* Structural basis for ligand recognition and activation of RAGE. *Structure* **18**, 1342-1352 (2010).
32 Zong, H. *et al.* Homodimerization is essential for the receptor for advanced glycation end products (RAGE)-mediated signal transduction. *J. Biol. Chem.* **285**, 23137-23146 (2010).
33 Penumutchu, S. R., Chou, R. H. & Yu, C. Structural insights into calcium-bound S100P and the V domain of the RAGE complex. *PloS one* **9**, e103947, doi:10.1371/journal.pone.0103947 (2014).
34 Mohan, S. K., Gupta, A. A. & Yu, C. Interaction of the S100A6 mutant (C3S) with the V domain of the receptor for advanced glycation end products (RAGE). *Biochem. Biophys. Res. Commun.* **434**, 328-333, doi:10.1016/j.bbrc.2013.03.049 (2013).
35 Delaglio, F. *et al.* NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6**, 277-293 (1995).
36 Garrett, D. S., Powers, R., Gronenborn, A. M. & Clore, G. M. A common sense approach to peak picking in two-, three-, and four-dimensional spectra using automatic computer analysis of contour diagrams. *Journal of Magnetic Resonance (1969)* **95**, 214-220 (1991).
37 Johnson, B. A. & Blevins, R. A. NMR View: A computer program for the visualization and analysis of NMR data. *J. Biomol. NMR* **4**, 603-614 (1994).
38 Otterbein, L. R., Kordowska, J., Witte-Hoffmann, C., Wang, C.-L. A. & Dominguez, R. Crystal structures of S100A6 in the Ca 2+-free and Ca 2+-bound states: the

calcium sensor mechanism of S100 proteins revealed at atomic resolution. *Structure* **10**, 557-567 (2002).
39 Cho, C. C., Chou, R. H. & Yu, C. Pentamidine blocks the interaction between mutant S100A5 and RAGE V domain and inhibits the RAGE signaling pathway. *Biochem. Biophys. Res. Commun.* (2016).
40 Barrangou, R. *et al.* CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**, 1709-1712 (2007).
41 Brouns, S. J. *et al.* Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* **321**, 960-964 (2008).
42 Garneau, J. E. *et al.* The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* **468**, 67-71 (2010).
43 Marraffini, L. A. & Sontheimer, E. J. CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *science* **322**, 1843-1845 (2008).
44 Edgar, R. & Qimron, U. The Escherichia coli CRISPR system protects from λ lysogenization, lysogens, and prophage induction. *J. Bacteriol.* **192**, 6291-6294 (2010).
45 Koonin, E. V., Makarova, K. S. & Zhang, F. Diversity, classification and evolution of CRISPR-Cas systems. *Curr. Opin. Microbiol.* **37**, 67-78 (2017).
46 Jiang, W. & Marraffini, L. A. CRISPR-Cas: new tools for genetic manipulations from bacterial immunity systems. *Annu. Rev. Microbiol.* **69**, 209-228 (2015).
47 Samson, J. E., Magadán, A. H., Sabri, M. & Moineau, S. Revenge of the phages: defeating bacterial defences. *Nature Reviews Microbiology* **11**, 675-687 (2013).
48 Pawluk, A., Davidson, A. R. & Maxwell, K. L. Anti-CRISPR: discovery, mechanism and function. *Nature Reviews Microbiology*, nrmicro. 2017.2120 (2017).
49 Rauch, B. J. *et al.* Inhibition of CRISPR-Cas9 with bacteriophage proteins. *Cell* **168**, 150-158. e110 (2017).
50 Dong, D. *et al.* Structural basis of CRISPR–SpyCas9 inhibition by an anti-CRISPR protein. *Nature* (2017).
51 Shing, J. *et al.* Disabling Cas9 by an anti-CRISPR DNA mimic. *bioRxiv*, 129627 (2017).
52 Yang, H. & Patel, D. J. Inhibition Mechanism of an Anti-CRISPR Suppressor AcrIIA4 Targeting SpyCas9. *Mol. Cell* (2017).
53 Bondy-Denomy, J. *et al.* Multiple mechanisms for CRISPR-Cas inhibition by anti-CRISPR proteins. *Nature* **526**, 136-139 (2015).
54 Chowdhury, S. *et al.* Structure reveals mechanisms of viral suppressors that intercept a CRISPR RNA-guided surveillance complex. *Cell* **169**, 47-57. e11 (2017).
55 Maxwell, K. L. *et al.* The solution structure of an anti-CRISPR protein. *Nature communications* **7** (2016).
56 Harrington, L. B. *et al.* A broad-spectrum inhibitor of CRISPR-Cas9. *Cell* **170**, 1224-1233. e1215 (2017).
57 Guillerez, J., Lopez, P. J., Proux, F., Launay, H. & Dreyfus, M. A mutation in T7 RNA polymerase that facilitates promoter clearance. *Proceedings of the National Academy of Sciences* **102**, 5958-5963 (2005).
58 Wagner, G. *et al.* Protein structures in solution by nuclear magnetic resonance and distance geometry: the polypeptide fold of the basic pancreatic trypsin inhibitor determined using two different algorithms, DISGEO and DISMAN. *J. Mol. Biol.* **196**, 611-639 (1987).
59 Bax, A., Kontaxis, G. & Tjandra, N. [8] Dipolar couplings in macromolecular structure determination. (2001).
60 Farrow, N. A., Zhang, O., Forman-Kay, J. D. & Kay, L. E. A heteronuclear correlation experiment for simultaneous determination of 15 N longitudinal decay and chemical exchange rates of systems in slow equilibrium. *J. Biomol. NMR* **4**,

727-734 (1994).
61 Tjandra, N., Wingfield, P., Stahl, S. & Bax, A. Anisotropic rotational diffusion of perdeuterated HIV protease from 15 N NMR relaxation measurements at two magnetic fields. *J. Biomol. NMR* **8**, 273-284 (1996).
62 Davis, D., Perlman, M. & London, R. Direct measurements of the dissociation-rate constant for inhibitor-enzyme complexes via the T1ρ and T2 (CPMG) methods. *Journal of Magnetic Resonance, Series B* **104**, 266-275 (1994).
63 Shen, Y., Delaglio, F., Cornilescu, G. & Bax, A. TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J. Biomol. NMR* **44**, 213-223 (2009).
64 Schwieters, C. D., Kuszewski, J. J. & Clore, G. M. Using Xplor–NIH for NMR molecular structure determination. *Progress in Nuclear Magnetic Resonance Spectroscopy* **48**, 47-62 (2006).
65 Nilges, M., Gronenborn, A. M., Brünger, A. T. & Clore, G. M. Determination of three-dimensional structures of proteins by simulated annealing with interproton distance restraints. Application to crambin, potato carboxypeptidase inhibitor and barley serine proteinase inhibitor 2. *Protein Engineering* **2**, 27-38 (1988).
66 Clore, G. M., Gronenborn, A. M. & Tjandra, N. (Elsevier, 1998).
67 Kuszewski, J., Qin, J., Gronenborn, A. M. & Clore, G. M. The impact of direct refinement against 13 C α and 13 C β chemical shifts on protein structure determination by NMR. *Journal of Magnetic Resonance, Series B* **106**, 92-96 (1995).
68 Clore, G. M. *et al.* The three-dimensional structure of α1-purothionin in solution: combined use of nuclear magnetic resonance, distance geometry and restrained molecular dynamics. *The EMBO journal* **5**, 2729 (1986).
69 Kuszewski, J., Gronenborn, A. M. & Clore, G. M. Improving the packing and accuracy of NMR structures with a pseudopotential for the radius of gyration. *J. Am. Chem. Soc.* **121**, 2337-2338 (1999).
70 Schwieters, C. D. & Clore, G. M. The VMD-XPLOR visualization package for NMR structure refinement. *J. Magn. Reson.* **149**, 239-244 (2001).
71 Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of applied crystallography* **26**, 283-291 (1993).
72 Kalodimos, C. G. *et al.* Structure and flexibility adaptation in nonspecific and specific protein-DNA complexes. *Science* **305**, 386-389 (2004).
73 Lee, G. M. *et al.* The structural and dynamic basis of Ets-1 DNA binding autoinhibition. *J. Biol. Chem.* **280**, 7088-7099 (2005).
74 Bhabha, G. *et al.* A dynamic knockout reveals that conformational fluctuations influence the chemical step of enzyme catalysis. *Science* **332**, 234-238 (2011).
75 Kovermann, M., Rogne, P. & Wolf-Watz, M. Protein dynamics and function from solution state NMR spectroscopy. *Q. Rev. Biophys.* **49**, e6 (2016).
76 Jinek, M. *et al.* Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science* **343**, 1247997 (2014).
77 Bondy-Denomy, J., Pawluk, A., Maxwell, K. L. & Davidson, A. R. Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system. *Nature* **493**, 429-432 (2013).
78 Fonfara, I. *et al.* Phylogeny of Cas9 determines functional exchangeability of dual-RNA and Cas9 among orthologous type II CRISPR-Cas systems. *Nucleic Acids Res.* **42**, 2577-2590 (2013).

IV. 국문초록

핵자기 공명 분광학을 이용한 S100A5-RAGE 복합체와 anti-CRISPR AcrIIA4의 구조 및 동력학연구

서울대학교 대학원

농생명공학부 바이오모듈레이션 전공

김익태

단백질 상호작용을 구조적관점에서 이해하는 것은 그 구조가 생물학적 기능과 밀접한 관련이 있기 때문에 중요하다. 본 연구에서는 핵자기 공명 분광학 및 열량측정기와 크기배제크로마토그래피 등의 생물리학적 방법을 이용하여 단백질 상호작용을 연구하였다.

1장에서는 칼슘결합 단백질 S100A5의 칼슘결합특성을 분석하고 최종당화산물수용체 (RAGE)와 RAGE의 기질, S100A5의 상호작용의 결합특성을 연구하였다. 열역학 분석을 통하여 S100A5는 칼슘을 두개의 C말단과 N말단의 EF-hand motif를 이용하여 각각 다른 친화력을 가지고 순차적으로 결합한다. 또한 이량체인 S100A5는 칼슘의존적으로 RAGE의 V 도메인과 결합하여 이종삼량체를 형성하

는 것을 화학량론적으로 밝혔다. 그리고 핵자기공명 적정 실험을 통해 S100A5 내에 이량체를 형성하는 결합부위가 RAGE와의 결합 부위에도 이용된다는 것을 밝혔다. 이러한 특이적인 결합방식과 화학량론은 RAGE에 의한 다양한 세포전달 신호를 제어를 이해하는데 도움이 될 것이다.

2장에서는 화농성연쇄상구균 (*Streptococcus pyogenes*)의 CRISPR-Cas9 단백질 (*Spy*Cas9)을 억제하는 anti-CRISPR AcrIIA4 단백질의 수용성 구조 및 동력학을 규명하였다. 본 연구에서 AcrIIA4는 세개의 알파헬릭스가 세개의 역평행 베타가닥에 얹혀져 있는 구조임을 밝혔다. 동력학 분석을 통하여 AcrIIA4의 *Spy*Cas9과의 결합부위인 고리 지역에 다양한 모션이 존재함을 밝혔다. 이러한 구조 및 동력학 특성이 *Spy*Cas9을 인식하는데 기여하는 것으로 추측한다. 또한 AcrIIA4는 *Spy*Cas9의 구조적 변화를 일으켜 특정 서열을 가진 목표 DNA의 인식을 유도하는 가이드 RNA 없이도 *Spy*Cas9과 상호작용함을 밝혔다. 이를 통해 박테리아 내에서 AcrIIA4와 *Spy*Cas9의 이종복합체가 형성되었다가 전사된 가이드 RNA와 결합하여 강한 결합의 삼량체가 되는 '순차적 결합 모델'을 제시하였다. 이러한 순차적 결합 모델에서는 AcrIIA4가 결합 목표 DNA와 경쟁하지 않기 때문에 박테리아 내에서 SpyCas9을 효과적

으로 억제할 수 있다. 본 연구가 CRISPR-Cas시스템을 이용한 유전체 편집 기술 응용 연구를 확장할 것으로 기대한다.

주요어: 핵자기공명분광학, 구조 및 동력학, AcrIIA4, Anti-CRISPR, Cas9, RAGE, S100A5